

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_nzp model

Run on: July 18, 2003, 16:45:27 ; Search time 34.4741 Seconds

(Without alignments)

6126.272 Million cell updates/sec

Title: US-09-806-276a-4

Perfect score: 4500

Sequence: 1 cggggctagccgagagacc.....cagctanaacgagggcagta 2546

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 segs, 41476328 residues.

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_nzp.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09806276/runat\_18072003\_160924\_12742/app.query.fasta\_1.3854  
-DB=SwissProt.40 -OPM=fastan -SUFFIX=isp -MINMATCH=0.1 -LOOCL=0 -LOOEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-USBR=050806276@cgn.1.1.79 @runat.18072003.160924\_12742 -NCPU=6 -ICPU=3  
-NO\_MMAP -LANG=ENGLISH -NEG.SCORES=0 -NAIT -DSFBLCK=100 -LONLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt.40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the total score distribution,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	222.5	4.9	356	1	BOR4_HUMAN
2	221	4.9	349	1	BOR4_MOUSE
3	195	4.3	391	1	MSEB_HUMAN
4	131	2.9	561	1	3BP2_HUMAN
5	130.5	2.9	741	1	GTSE_MOUSE
6	122.5	2.7	3530	1	MY15_HUMAN
7	120.5	2.6	519	1	CBX2_MOUSE
8	115.5	2.6	982	1	HB11_MOUSE
9	115	2.6	1475	1	N153_HUMAN
10	115	2.6	1833	1	ZEP2_HUMAN
11	111.5	2.5	596	1	YAC1_MOUSE
12	110.5	2.5	1117	1	KDZ2_HUMAN
13	110	2.4	501	1	OPFX_HUMAN
14	110	2.4	628	1	V70K_TYRVA
15	110	2.4	2351	1	FAB_HUMAN
16	109	2.4	627	1	TSK_MOUSE
17	108.5	2.4	889	1	CTKF_RAT
18	108	2.4	601	1	3BP1_MOUSE

19	107.5	2.4	528	1	POD2_HUMAN	000592 homo sapien
20	107.5	2.4	2715	1	TRX2_HUMAN	09um6 homo sapien
21	107	2.4	503	1	WAP_HUMAN	043516 homo sapien
22	107	2.4	1128	1	BEM3_YEAST	P32873 saccharomyc
23	106	2.4	732	1	TAU_MOUSE	P10637 mus musculu
24	105.5	2.3	751	1	TAU_RAT	P19332 mus musculu
25	105.5	2.3	3669	1	HRX_HUMAN	P19332 mus musculu
26	105	2.3	920	1	AD19_MOUSE	003164 homo sapien
27	104.5	2.3	551	1	TFE5_HUMAN	035674 mus musculu
28	104.5	2.3	864	1	WS14_MOUSE	004206 mus musculu
29	104.5	2.3	1805	1	RMI_HUMAN	09um23 mus musculu
30	104.5	2.3	3163	1	POLG_TYRVA	092545 homo sapien
31	104	2.3	2805	1	MAPA_HUMAN	002597 t genome po
32	104	2.3	842	1	PROP_MOUSE	P53235 saccharomyc
33	103.5	2.3	437	1	GRSE_HUMAN	P78559 mus musculu
34	103.5	2.3	720	1	ZNA2_HUMAN	P1680 mus musculu
35	103.5	2.3	734	1	PCGN_MOUSE	09um23 mus musculu
36	103.5	2.3	1268	1	SG22_DROME	P28698 homo sapien
37	103.5	2.3	1365	1	AREA_GIBFU	P55066 mus musculu
38	103	2.3	870	1	DPOL_HPEVA	P25172 drosophila
39	103	2.3	971	1	SBP2_HUMAN	060309 homo sapien
40	102.5	2.3	832	1	TSH_DROME	P78688 gibberella
41	102.5	2.3	854	1	TPSN_HUMAN	P24024 hepatitis b
42	102.5	2.3	993	1	YHL1_EBV	096f21 homo sapien
43	102	2.3	448	1	ICP0_HSV1	P22265 drosophila
44	102	2.3	660	1		015533 homo sapien
45	102	2.3	775	1		P03181 epstein-bar
						P08393 herpes simp

# ALIGNMENTS

RESULT 1  
ID BOR4\_HUMAN  
AC 098301: 095828: 096FT3: PRT: 356 AA.  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE BLinder of Rho GTPase 4 (Cdc42 effector protein 4).  
GN BOR4 OR CEP4.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ileal mucosa;  
RX MEDLINE=21036164; PubMed=1185749;  
RA Osada N., Kusuda J., Suzuki Y., Sugano S., Hashimoto K.;  
RT "Sequence analysis, gene expression, and chromosomal assignment of  
RT mouse Bor4 gene and its human orthologue.";  
RL J. Hum. Genet. 45:374-377(2000).  
RN [2]  
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.  
RC TISSUE=Eye;  
RX MEDLINE=21125609; PubMed=11035016;  
RA Hirsch D.S., Pirone D.M., Burbeo P.D.;  
RT "A new family of cdc42 effector proteins, CEPs, function in fibroblast  
RT and epithelial cell shape changes.";  
RL J. Biol. Chem. 276:875-883(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX Strausberg R.;  
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP cytoskeleton. May act downstream of CDC42 to induce actin  
RN [5]  
RP pseudopodia formation, when overexpressed in fibroblasts.  
RN [6]  
RP SUBUNIT: Interacts with CDC42 and TCO1, in a GTP-dependent manner  
RN [7]  
RP (by similarity).  
RN [8]  
RP SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated.  
RN [9]  
RP TISSUE SPECIFICITY: Not detected in any of the adult tissues

```

Db      141 SerSerProVal-----LYSLYSALAAsnAspGlyGlyGly----- 152
QY      905 GTTCACACGAGGAGNACACTGTGGGCTCCAGCGGTTGTGCATTCATGCAGCCAGCCAGGC 964
Db      153 -----GLYAspGluGluValaGlyPThrGluGluValaValaProArgArgAsnGlyAla 169
QY      965 AGAGACAGCCACTCTCCAGC-----CTGTCCGACACAGTACCCGACACGCGAGCCGAG 1018
Db      170 ALAGLYProHisSerProAspProLeuAspGluVala-----AlaPheGly 185
QY      1019 GACATGTTTGACACCAATCCCAACCCAGTCAGCTCATCAAGGAAAAAGACTAGTCAGAGCAG 1078
Db      186 AsperThrAspLeuProValValaProLysAlaIaThrGlyGlyLeuLysAlaGlu--- 204
QY      1079 TCCCTCTGTGCACCTTACAGGTTTCCTCTCTCTCTCTCTGACAGCTATGATTTGGGCGCTACTT 1138
Db      205 -----SerIleMetSerPheHisIleAspLeuGlyProSerMet 217
QY      1139 TTGATATAGGTGCTGAATGATGAATGATAAATAATAG 1174
Db      218 LeuGlyAspValLeuSerIleMetAspLysGluGlu 229

RESULT 2
BOR4_MOUSE
ID      BOR4_MOUSE      STANDARD:      PRT:      349 AA.
AC      09JUN96: 090278:
DT      15-JUN-2002 (Rel. 41, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Binder of Rho GTPase 4 (Cdc42 effector protein 4).
      BOR4 OR CEP4.
      Mus. musculus (Mouse).
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
      NCBI_TaxID=10090;
      [1]
      SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
      PC      TISSUE=ileal mucosa;
      RX      MEDLINE=21036164; PubMed=11185749;
      RA      Oada N., Kusuda J., Suzuki Y., Sugano S., Hashimoto K.;
      RT      "Sequence analysis, gene expression, and chromosomal assignment of
      RL      mouse Borq4 gene and its human orthologue.";
      J. Hum. Genet. 45:374-377(2000).
      [2]
      SEQUENCE FROM N.A.
      RA      Strausberg R.;
      RL      Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
      [3]
      SEQUENCE OF 1-143 FROM N.A., AND INTERACTION WITH TC10 AND CDC42.
      RP      MEDLINE=99421943; PubMed=10490598;
      RX      Joberty G., Perlungher R.R., Macara T.G.;
      RT      "The borgs, a new family of Cdc42 and TC10 GTPase-interacting
      RL      proteins.";
      Mol. Cell. Biol. 19:6585-6597(1999).
      [4]
      FUNCTION: Probably involved in the organization of the actin
      cytoskeleton. May act downstream of CDC42 to induce actin
      filament assembly leading to cell shape changes. Induces
      pseudopodia formation, when overexpressed in fibroblasts.
      [5]
      SUBUNIT: Interacts with CDC42 and TC10, in a GTP-dependent manner.
      [6]
      SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated (by
      similarity)
      [7]
      TISSUE SPECIFICITY: Ubiquitous.
      [8]
      SIMILARITY: BELONGS TO THE BOR4/CEP FAMILY.
      [9]
      SIMILARITY: CONTAINS 1 CTRB DOMAIN.
      [10]
      This SWISS-PROT entry is copyright. It is produced through a collaboration
      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
      the European Bioinformatics Institute. There are no restrictions on its
      use by non-profit institutions as long as its content is in no way
      modified and this statement is not removed. Usage by and for commercial
      entities requires a license agreement (See http://www.isb-sib.ch/announce/
      or send an email to license@isb-sib.ch).

```

Db	196	-----	SettleUsenSerPhenHis1LAspLeuglyProSerMetLeuglyAspValLeu	213
Oy	1154	AATGTAATGATATAAATGATG	1174	
Db	214	SettleMetasplysAspScin	220	
RESULT 3				
MSES5_HUMAN		STANDARD:	PRT:	391 AA.
AC	000587; Q96GN1			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Serum protein MSES5.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI:TaxId=9606;			
NP	1			
RP	SEQUENCE FROM N.A. (ISOFORM 1)			
RX	MEDLINE=20057165; PubMed=10591208;			
RA	Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,			
RA	Clamp M., Smink L.J., Alnough R., Almeida J.P., Babbage A.K.,			
RA	Baggaley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,			
RA	Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,			
RA	Burrill W.D., Burton J., Carter C., Carter N.P., Chen Y., Clark G.,			
RA	Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,			
RA	Conroy D., Corby N.R., Coville S.J., Cox A.V., Davis J., Dawson E.,			
RA	Dhant P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,			
RA	Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,			
RA	Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,			
RA	Hall R.E., Hall-Tamlyn G., Heathcote R.N., Ho S., Holmes S.,			
RA	Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,			
RA	Laird G.K., Langford C.F., Leverisa M.A., Lloyd C., Lloyd D.M.,			
RA	Matlin I.D., Mashregh-Mohammadi M., Matthews L.H., Mccann O.T.,			
RA	Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,			
RA	Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,			
RA	Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,			
RA	Scott C.E., Sehra H.K., Skuce C.B., Smalley S., Smith M.L.,			
RA	Soderlund C., Spragon L., Stewart C.A., Sulston J.E., Swann R.M.,			
RA	Vardin M., Wall M., Wallis J.W., Whiteley M.N., Willey D.L.,			
RA	Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,			
RA	Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,			
RA	Mimosima S., Kawasaki K., Sasaki T., Asakawa S., Kidoh J.,			
RA	Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,			
RA	Roe B.A., Chen F., Chu L., Cradtree J., Deschamps S., Do A., Do T.,			
RA	Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.,			
RA	Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,			
RA	Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,			
RA	Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,			
RA	Fulton R., Johnson D., Chissoe S., Murray J., Miller N., Minx P.,			
RA	Fulton R., Johnson D., Bemis G., Bentley J., Brdshaw H., Bourne S.,			
RA	Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,			
RA	Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,			
RA	Schee P., Walker C., Wamsley A., Waymann P., Pepin K., Nelson J.,			
RA	Korfi I., Beceji J.A., Hillier L., Mardis E., Waterston R., Wilson R.,			
RA	Manuel B.S., Shalk T., Kurahashi H., Saito S., Budarf M.L.,			
RA	Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,			
RA	Kim U.J., Shinya H., Simon M.I., Dumanski J.P., Peyrard M., Kedia D.,			
RA	Seroussi E., Franssen I., Tapia I., Brider C.E., O'Brien K.P.,			
RA	Wilkinson P., Bodenecini A., Hartman K., Hu X., Khan A.S., Lane L.,			
RA	Tilman Y., Wright H.,			
RT	The DNA sequence of human chromosome 22. ",			
RL	Nature 402:489-495(1999).			

RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Pancreas;  
 RA Strausberg R.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be  
 CC produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: ENDOTHELIAL AND BONE MARROW STROMAL CELLS.  
 CC -1- SIMILARITY: CONTAINS 1 CRIB DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: M88338; AAA36606.1;  
 DR EMBL: AL022315; CAB42833.1;  
 DR EMBL: BC009356; AAH09356.1;  
 DR PIR: A42973; A42973.  
 DR MIM: 606084;  
 DR InterPro: IPR000095; PAKbox/Rho-binding.  
 DR SMART: SM00285; PBD; 1.  
 DR PROSITE: PS50108; CRIB; 1.  
 DR Repeat: Alternative splicing.  
 FT DOMAIN 38 52  
 FT DOMAIN 220 275  
 FT VAAPLIC 258 264 MISSING (IN ISOFORM 2).  
 SQ SEQUENCE 391 AA; 40294 MW; 625ER3C11P20C2B CRC64;  
  
 Alignment Scores:  
 Pred. No.: 1.6e-07 Length: 391  
 Score: 195.00 Matches: 74  
 Percent Similarity: 42.51% Conservative: 31  
 Best Local Similarity: 29.96% Mismatches: 76  
 Query Match: 4.33% Indels: 66  
 DB: 1 Gaps: 9  
  
 US-09-806-276a-4 (1-2546) x MSE5\_HUMAN (1-391)  
 QY 449 AATAACAGAAAGAAAGAAATTAACAGGAGACATTCCTGCTATATGATGACAT 508  
 DB 25 SerSerSerGlnGlyLysArg-----ArgLeuThrAlaAspMetLieser 39  
 QY 509 CCCCCTGGAGAGATTTCGCCACACCATCCACATTGGCAAGAGGCCAGCAGATGTC 568  
 DB 40 HisProLeuGlyAspRheArgHisThrMetHisValGlyArgGlyLys-----AspVal 57  
 QY 569 TTGGAGATATTTCCTTCTT-----CAAGGAACACTACAGACTTTTA 610  
 DB 58 PheGlyAspThrSerPheLeuSerAsnHisGlyLysSerGlySerThrHisAspSer 77  
 QY 611 CCGTGAAGC-----CAGGAGAAAGACACACCTGGGCGATTCCTGGGCAATAATGATTC 664  
 DB 78 ProArgSerPheLeuAlaLysLysLeuGlnLeuValAlaArgValGlyAlaProProArg 97  
 QY 665 TTCGGGCCCAACAGACCTCGACTGTGTTCACAGAAAGCCCTCCCGGTGCCTCAA 724  
 DB 98 ArgMetAlaSerProProAlaProSerProAlaProProAlaLieserProLieser 117  
 QY 725 AATGCCATCCCTCCGACCATTTGGAGAGATCCCAAGCTCATGTTGCCCTATTGTCA 784  
 DB 118 AsnAlaLieserLeuProGlnLeu-----125  
 QY 785 CCAAGTACATTAAATTCACACAGAGCTCTCGGGCCAGCAAAAGTCCAGGCTTAGC 844  
 DB 126 -----AsnGlnAlaAlaLysAspSerLeuValValGlyLysLeuSer 139  
 QY 845 TCGGAGCCCGTCATGGAGAAAGAAAGCTCAGAGAAAGAAAGCACTGCTGGAGAAATGGGACA 904  
 DB 845 TCGGAGCCCGTCATGGAGAAAGAAAGCTCAGAGAAAGAAAGCACTGCTGGAGAAATGGGACA 904

DB 140 PheAsp-----141  
 QY 905 GTCCACACGAGGAGACACCTGCTGGGCTCCAGCGCTTCGATCATGTCACGCAAGGC 964  
 DB 142 -----SerSerProThrSerSerThrAspGlyHisSerSerTyrGly 155  
 QY 965 AGAGACAGCCACCTCCAGCGCTTCGCA-----CAGTACCAGCATGCGCAGCGAGAC 1021  
 DB 156 LeuAspSerGlyPheCysThrLieserArgLeuProArgSerGlyLysProHisAspArg 175  
 QY 1022 ATGTTGACCATCCACCCATCGAGCTCATCAAGGAAAGACTAGTCAGAGAGTCC 1081  
 DB 176 AspArgAspGlySerPheProSerGln-----ProGlyLeuArgArgSerAspSerLeu 193  
 QY 1082 CTCCTGACCTTACAGAGTTCCCTCCTCCGACGCTGATCTTGGGCGCTGACTTTTG 1141  
 DB 194 LeuSer-----PheArgLeuAspLeuAspLeuGlyProSerLeu 207  
 QY 1142 GATGAGTGTGATGATGATG 1162  
 DB 208 SerGlnLeuLeuGlyValMet 214  
  
 RESULT 4  
 ID 3BP2\_HUMAN STANDARD; PRT; 561 AA.  
 AC P78314; O15373; O00500; P78315;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE SH3BP2, an SH3 binding domain protein 2 (3BP-2).  
 GN SH3BP2 OR 3BP2 OR RESA-23.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 RX [1]  
 RN SEQUENCE FROM N.A. (LONG ISOFORM).  
 RP MEDLINE=97446134; PubMed=9299232;  
 RA Bell S.M., Shaw M., Jou Y.-S., Myers R.M., Knowles M.A.;  
 RT "Identification and characterization of the human homologue of  
 RT SH3BP2, an SH3 binding domain protein within a common region of  
 RT deletion at 4p16.3 involved in bladder cancer.";  
 RL Genomics 44:163-170(1997).  
 RN [2]  
 RN SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS), AND TISSUE SPECIFICITY.  
 RP TISSUE=Brain;  
 RX MEDLINE=98403881; PubMed=9734812;  
 RA Hadano S., Ishida Y., Ikeda J.-E.;  
 RT "The primary structure and genomic organization of five novel  
 RT transcripts located close to the Huntington's disease gene on human  
 RT chromosome 4p16.3.";  
 RL DNA Res. 5:177-186(1998).  
 RN [3]  
 RN SEQUENCE FROM N.A. (LONG ISOFORM).  
 RP TISSUE=Tonsil;  
 RA Gokmeijer J., Delgiannidis K.E., Ligris K., Ernst T.J.;  
 RT "3BP2 binds to phosphatidylinositols, linking the hemopoietic tyrosine  
 RT kinase c-FES to the cytoplasmic membrane in a phosphorylation  
 RT dependent mechanism.";  
 RL Blood 88:473A-473A(1996).  
 RN [4]  
 RN VARIANTs CRBM GLN-415; PRO-415; ARG-418; HIS-418; LEU-418; ARG-420 AND  
 RP GLU-420.  
 RX MEDLINE=21275962; PubMed=11381255;  
 RA Ueki Y., Tiziani V., Santanna C., Fukui N., Maulik C., Garfinkle J.,  
 RA Niimiya C., domarai C., Peters H., Habel M., Rhee-Morris L.,  
 RA Doss J.B., Kreiborg S., Olsen B.R., Reichenberger E.;  
 RT "Mutations in the gene encoding c-Abl-binding protein SH3BP2 cause  
 RT cherubism.";  
 RL Nat. Genet. 28:125-126(2001).  
 CC -1- FUNCTION: BINDS DIFFERENTIALLY TO THE SH3 DOMAINS OF CERTAIN  
 CC PROTEINS OF SIGNAL TRANSDUCTION PATHWAYS. BINDS TO  
 CC PHOSPHATIDYLINOSITOLS; LINKING THE HEMOPOIETIC TYROSINE KINASE



Query Match:	2.91%	Indels:	99
DB:	1	Gaps:	13
US-09-806-276A-4 (1-2546) x 3BP2_HUMAN (1-561)			
QY 245 ACCTGAGGAGGCTCTTTGGACCCGACATTCATTTCGAAGCAGCTTGTGAAGCAGCTTGA	30		
Db 142 ThSerAspSerSerSerAspThrAspSerPheThrIylAlaValGluArgProVal	16		
QY 305 GAAAATGATGTGACAGTT---CCATCAAAAAGAGATTCAAGAAATATACATCTGTGA	36		
Db 161 -----AspIleSerLeuSerProTyrProThrAspAsnGluAspTyrGluHisAspAs	17		
QY 362 AGAAGTGGCCCTTTC-----TCCCGTTCCAAAATAGACA	39		
Db 178 peltuAspAspSerTyrLeuGluProAspSerProGluProGlyArgLeuGluAspAla	19		
QY 398 TTCTCAAAATTCGAAAATGCCAGCCAGAACCCCAATTACTGTGAAGCAGCCAAATAACAG	45		
Db 198 uMethIisProProAlaTyrProProProProValProThrProArgLysProAlaPheSe	21		
QY 458 AAGAAGAAATTTAAACTGAGGACATTTCTGCTCTGATGATGATCAGTCCCGCTT	51		
Db 218 r-----AspMetProArgAl	22		
QY 518 GGAGACTTTCGCCACACACCATTCACATGTCGCAAGAGGCGACAGCATCTTTGGAGAT	57		
Db 223 ahIserPheThrSer-----	22		
QY 578 ATTCTCTTTCGAAGGAGACACAGACTTTTACTCGAAACAGAGAGAACACACACTG	63		
Db 229 -----LysGlyProGluProLeuLeuProProProProPolysSHs	61		
QY 638 GGCCAGTTCCTCGGCGCAATTAATGAGTTCTTCGGGCGCCACAGAGACCTCGGACTGTGTTC	69		
Db 243 yLeuProAspValGlyLeuAlaIaIaGluAspSerLysArgAspPro-----LeuGlySer	261		
QY 658 ACAGAAAGCCCTCCCGGTGCTCAAAAAATGCCATTCCTCCGACCATTTGAGAGATCC	75		
Db 261 oArgArgAlaGluPro-----CysProArgValProAlaThrProAlaGlyMetSerAspPr	28		
QY 758 CACAGTCTCATGTTGCCCTATTGTGCACAGTGACATTTAATTCGAAACGAGACTCTTC	81		
Db 280 oProLeuSerThrMetPro-----	28		
QY 818 GGGCCAGCAAGCTGCCAGAGGTCTTAGCTCGAGCCGTCATGAGAGAAAGCTCAGAG	87		
Db 287 -----ThrAlaProGlyLeuAlaArgLysProProCysPheArgGluSerAlaSerPr	303		
QY 878 AAAAGCAGTCTGTGGAGATGGGACAGTCCACAGGAGACACTGATGGGCTCCAGC	93		
Db 303 oSerProGluProTyrThrProGlyHisGlyAlaCysSerThrSerSerAlaAlaIleMe	323		
QY 938 GGT-----CTGATCTCAGTCCAGGCAA	961		
Db 323 tAlaThrAlaThrSerArgAsnCysAspLysLeuLysSerPheHisLeuSerProArgI	343		
QY 962 GCGAGAGACAGCCATCTCTCAGACCTGTCCGAAAGATACCCCG-----	100		
Db 343 yProProThrSerGluProProProValProAlaAsnLysProLysPheLeuLysIleAl	363		
QY 1005 ----ACTGGCAGCCGAGGACATG-----TTGACATCTCCACCCCATCGCAG	1041		
Db 363 agLIGluAspProProAlaArgGluAlaAlaMetProGlyLeuPheValProPro-----	380		
QY 1049 CTCATCAAGGAGAAAGCTAAGTCAAGAGAGAGCCCTCTCAACCTTACAGTGTCCCTCC	110		
Db 381 -----ValAlaProArgProProAlaLeuLysLeuProValProGluAl	395		
QY 1109 T 1109			
Db 395 a 395			

RESULT 5  
ID GTSE\_MOUSE STANDARD: PRT: 741 AA.  
AC Q8R080; Q8R015; Q9CS69;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE G2 and S phase expressed protein 1 (Gtse-1) (899 protein).  
GN GTSE1 OR B99.  
OS Mus musculus (Mouse)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98393542; PubMed=9724637;  
RA Ureña R., Collavin L., Lazarevic D., Delia D., Schneider C.;  
RT "A novel p53-inducible gene coding for a microtubule-localized protein  
with G2-phase-specific expression.";  
RL EMO J. 17:5015-5025(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Memory gland;  
RA Strausberg R.;  
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-130 FROM N.A.  
RX STRAIN=C57BL/6J; TISSUE=Embryo;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Akawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gliss C., King B., Kochiya H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schirral L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Botfield D., Boujarda N., Cantinelli P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Guarnicini S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [4]  
RP DEVELOPMENTAL STAGE AND PHOSPHORYLATION.  
RX PubMed=10984615;  
RA Collavin L., Monte M., Verardo R., Pflieger C., Schneider C.;  
RT "Cell-cycle regulation of the p53-inducible gene B99.";  
RL FEBS Lett. 481:57-62(2000).  
CC -i- FUNCTION: May be involved in p53-induced cell cycle arrest in G2/M  
phase by interfering with microtubule rearrangements that are  
required to enter mitosis. Overexpression delays G2/M phase  
progression.  
CC -i- SUBCELLULAR LOCATION: Cytoplasmic. Associated with microtubules.  
CC -i- DEVELOPMENTAL STAGE: Expression begins at S phase, accumulates in  
late S/G2 phase and disappears in G1 phase.  
CC -i- PTM: Phosphorylated in mitosis.  
CC -i- INDUCTION: By p53 when exposed to different DNA damaging  
agents, including gamma irradiation and chemotherapeutic drugs.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/annouce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL: AJ222580; CA10848.1; -  
DR EMBL: BC027213; AAH27213.1; -  
DR EMBL: AK012870; BAB28524.1; -  
DR MGD: MGI:135275; Gtse1.  
KW Microtubules; Phosphorylation.  
FT DOMAIN 23 28 POLY-SER.  
FT CONFLICT 153 153 S -> P (IN REF. 2).  
FT CONFLICT 246 246 I -> L (IN REF. 2).  
FT CONFLICT 252 252 V -> M (IN REF. 2).  
FT CONFLICT 274 274 S -> G (IN REF. 2).  
FT CONFLICT 398 398 SP -> IL (IN REF. 2).  
SQ SEQUENCE 741 AA; 78751 MW; 10684D06EE9B7ACB CRC64;  
Alignment Scores:  
Pred. No.: 0.0221 Length: 741  
Score: 130.50 Matches: 111  
Percent Similarity: 32.98% Conservative: 43  
Best Local Similarity: 23.77% Mismatches: 145  
Query Match: 2.90% Indels: 168  
DB: Gaps: 22  
US-09-806-276a-4 (1-2546) x GTSE\_MOUSE (1-741)  
QY 3 GGGCTAGCCCGGAGACCGCCGCGCTGGGCGCTTCACGCCCTTCGAGCGGAT 62  
||| |||:  
DB 330 GilylAlaIySerSerSerSerSerSerSerSerSerSerSerSerSerSer 347  
63 AATGCGGTAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 122  
||| |||:  
DB 348 ArgArgIhSerThrSerSerSerSerSerSerSerSerSerSerSerSerSer 356  
123 TTS---CGTGGCGCGCGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGGT 179  
||| |||:  
DB 367 LeuProAlaAlaProAlaAlaValPheGlyArgAlaAlaAlaAlaAlaAlaAla 385  
180 TTGTCTGT 239  
385 -----  
DB 240 CCGTAACCTCAGAGAGGCTCTCTTACCCCTCATCTCATTTGAGAGCCACTTGAAGCA 299  
||| |||:  
DB 386 -----GlnThrValAlaGluGlnProLysVal-----ProThrLeuSerPro 399  
300 CTGAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 359  
||| |||:  
DB 400 LeuThrGlnGlnProGlnThrPro-----  
360 GAGAAAGTGGCGCGCTTTCGCCCTTGCAGAAATAGACATTCGAAATTCGAAATGCCAG 419  
||| |||:  
DB 408 GlnGlnArgGlyProArgLeuAspProAspThrGlnThrProGlnLeuAsnLysThrVal 427  
420 CCAAGACCCCAATTTACTGTAAGAGCAGCAATTAACAGAAAGAAATTTAACTGA 479  
427 -----  
DB 480 GGGCATTTCTGTCTCTGATATGATGATGATGATGATGATGATGATGATGATGAT 539  
428 -----SerIleLysAlaArgAspSerIleLysSerCysLys 439  
540 ACATTGGCAAGAGAGCGCCAGCAGAGCTTTGGAGATTTCTTTCTTCAAG----- 593  
||| |||:  
DB 440 ThrGlnAlaValSerThrThrThrAspProPheLysVal-----ProGlnPheSerValGly 458  
594 -----GSACTACGAGCTTTTACCTGGAACAGGAGAAAGAGCAACACCTGGGCG 641  
459 GluSerProGlyGlyValThrProLysPheSerArgThrHisArgLysGlnSerThrPhe 478  
642 AGTTCCCTGGGCAATAGAGCTCT-----TCCGGGCAACAGCAGCAGCTCGG 686  
479 ProAlaSerArgValValSerSerThrProValArgArgSerSerGlyThrThrProGln 498  
687 ACTGTGTCTCAGCAAGAAAGCGCTGCC-----CGGTGTCAAAAATGCCATCT 734

RA Wang A., Jiang Y., Fidelel R.A., Probst F.J., Wilcox E.R.,  
RA Touchman J.W., Morton C.C., Morell R.J., Noben-Trauth K., Camper S.A.,  
RA Friedman T.B., "Association of unconventional myosin MYO15 mutations with human  
RT nonsyndromic deafness DFNB3."; *Science* 280:1447-1451(1998)  
RL Science 280:1447-1451(1998)  
CC -1- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE  
CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.  
CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS  
CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY  
CC SIMILARITY). MAY PLAY A ROLE IN THE FORMATION OR MAINTENANCE OF  
CC THE ACTIN-RICH STRUCTURES OF THE INNER EAR SENSORY HAIR CELLS.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PITUITARY, ALSO EXPRESSED  
CC AT LOWER LEVELS IN ADULT BRAIN, KIDNEY, LIVER, LUNG, PANCREAS,  
CC PITUITARY AND SKELETAL MUSCLE. NOT EXPRESSED IN BRAIN. IN THE  
CC DISEASE: Defects in MYO15A are the cause of an autosomal recessive  
CC form of nonsyndromic deafness (DFNB3)  
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
CC -1- SIMILARITY: CONTAINS 3 IQ DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC -----  
CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL, AF144094; AAF05903.1; -  
DR EMBL, AF051976; -; NOT\_ANNOTATED\_CDS.  
DR HSSP, P10587; 1BR2.  
DR Genew; HGNC:7594; MYO15A.  
DR MIM; 602666; -;  
DR MIM; 600316; -;  
DR InterPro; IPR000299; Band\_4.1.  
DR InterPro; IPR000048; IQ\_region.  
DR InterPro; IPR000857; MYTH4.  
DR InterPro; IPR001452; SH3.  
DR InterPro; IPR001609; myosin\_head.  
DR Pfam; PF00063; myosin\_head; 1.  
DR Pfam; PF00612; IQ\_3.  
DR Pfam; PF00784; MYTH4\_2.  
DR PRINTS; PRO0193; MYOSINHEAVY.  
DR PRODom; PDO00355; myosin\_head; 1.  
DR SMART; SM00015; IQ; 2.  
DR SMART; SM00242; MYSC; 1.  
DR SMART; SM00139; MYTH4; 2.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS00660; BAND\_41\_1; FALSE\_NEG.  
DR PROSITE; PS00661; BAND\_41\_2; FALSE\_NEG.  
DR PROSITE; PS50057; BAND\_41\_3; 1.  
DR PROSITE; PS50096; IQ; 3.  
DR PROSITE; PS50092; SH3; FALSE\_NEG.  
KW Myosin, ATP-binding, Actin-binding, Coiled coil, Repeat, SH3 domain;  
KW Calmodulin-binding, Disease mutating, Deathness.  
FT DOMAIN 1 1887  
FT FT 1888 2029 HEAD OR MOTOR DOMAIN.  
FT FT 2030 3530 NECK OR REGULATORY DOMAIN.  
FT FT 3531 3530 TAIL.  
FT DOMAIN 1323 1350 COILED COIL (POTENTIAL).  
FT DOMAIN 1792 1799 ACTIN-BINDING (POTENTIAL).  
FT DOMAIN 1902 1924 IQ 1.  
FT DOMAIN 1925 1954 IQ 2.  
FT DOMAIN 1955 1976 IQ 3.  
FT DOMAIN 2867 2953 SH3.  
FT NP\_BIND 3206 3443 BAND 4.1-LIKE.  
FT NP\_BIND 1315 1322 ATP (POTENTIAL).  
FT VARIANT 2111 2111 N -> Y (IN DFNB3; FAMILY FROM BENGKALA).  
FT VARIANT 2111 2111 /FTID=VAR\_010303.  
FT I -> F (IN DFNB3; INDIAN FAMILY).  
FT VARIANT 2113 2113

FT SEQUENCE 3530 AA: 395171 MW: 3D103923D4BCBE4A CRC64;  
/FT16-VAR\_010304.

Alignment Scores: 0.142 Length: 3530  
Pred. No.: 122.50 Matches: 70  
Score: 33.968 Conservative: 21  
Percent Similarity: 26.128 Mismatches: 108  
Best Local Similarity: 2.728 Indels: 69  
Query Match: 1 Gaps: 16

US-09-806-276a-4 (1-2546) x MY15\_HUMAN (1-3530)

```

OY 369 GGCCCTTTCCTCCGCTTGAATAATAGACATTCCTCAATTCGAATAATGCCAGCCAGACCC 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 759 GlylaserProAlaSerArgAlaSerProAlaSerPro----- 776
OY 429 CAATTACCTGAAAGCAGCCATTAACAGAAAGAAAGAAATTAACCTGAGGACATTC 488
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 777 -----GlnProSerLeuAlaArgSerProGlyLeuGly-----Tyr 788
OY 489 TGCTCTCGATGATGATCAGTCCCGCTTGAGACATTCGCCACACCATCCACATTCGCA 548
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 789 Cys----- 793
OY 549 AAGAGGGCCAGCAGATGCTTTGGAGATATTCCTTCTTCAAGGAACTACAGACTTT 608
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 794 ProProSerProGlnLeuSerLeuAlaArgThylGlyProhe-----GlnProProPhe 810
OY 609 TACTGTGAAGCAGAGAGAACACACTGGGCCAGTTCCCTGGGCATTAATGAGTTCCTCC 668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 811 LeuProProAlaArgArgProAlaArgSerLeuGlnGlnSer----- 823
OY 669 GGGCCAGCAGCAGCTCGGACCTGTGTTCACAGAAAGCCCTCCCGGCTCTCAAAATG 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 824 -----ProAlaProAlaArgAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 840
OY 729 CCATCTCCCTCCGACCATGAGAGATCCCAAGCTCATGTTGCGCTTATTGTACACAG 788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 841 ProGlySerProAlaPro-----ProSerProProLeuGlnGlnGlnGlnGlnGlnGlnGln 859
OY 789 TGACATTTA---ATTCCAAACAGAGATCCTTCGGGCCAGCAAGCTGC-----CCA 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 860 SerSerLeuAlaSerLeuProAlaArgLeuProAlaArgLeuSerGlnProPro 879
OY 837 GCGTTAGCTCGAGCCCTGCATGAGAGAAAGAGCTCAGAGA-----AAAGCACTGTGT 890
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 880 ThrArgAlaValAlaSerProGlnValAlaArgLeuProAlaArgProProAlaArgAla 899
OY 891 TGGGAGATGGGACAGTCCACAGAGAGACACTCGTGGGGCTCCAGCGTTCTGCATCTC 950
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 900 TrrArgAlaProLeuGlnAlaArgGlnSerProArg----- 911
OY 951 AGTCCAGCCAGCAGAGACA-----GCCACTCCCTCCAGCTGTCG----- 992
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 912 GlnProGlnAlaSerGlnThrProTrrPthValProProLeuAlaProSerTrrPspAl 931
OY 993 AAGCAGTACCCGAGCTGGCCAGCAGAGACATGTTTGACCATCCCA---CCCATGCGAGC 1049
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 932 AspMetProProThrGlnArgProProSer-----ProTrrProGlnGlnGlnGln 947
OY 1050 TCATCAGAGGAAGACTAGTACAGAGAGTCCCTCTCTGACCTTACAGGCTCCCTCTCT 1109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 948 GlySerArgArgGlyPheSerArgProProPro-----ValProGlnAlaSer 962
OY 1110 CCTTCAGCTGATCTGGGCTT 1133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 963 ProPheLeuGlnLeuLeuGlyPro 970

```

RESULT 7  
CBX2\_MOUSE STANDARD: PRT: 519 AA.  
ID CBX2\_MOUSE  
AC P30658; 035731.

```

DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chromobox protein homolog 2 (Modifier 3 protein) (M33).
GN CBX2 OR M33.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RC MEDLINE=92315938; PubMed=1352241;
RA Pearce J.J.H., Singh P.B., Gaunt S.J.;
RT "The mouse has a Polycomb-like chromobox gene.";
RT Development 114:921-930(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA Schoorlemmer J., Vidal M.;
RT "Ring1a is a transcriptional repressor that interacts with the
RT Polycomb-M33 protein and is expressed at rhombomere boundaries in the
RT mouse hindbrain.";
RT submitted (May-1997) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: INVOLVED IN MAINTAINING THE TRANSCRIPTIONALLY REPRESSIVE
CC STATE OF GENES. MODIFIES CHROMATIN, RENDERING IT HERITABLY CHANGED
CC IN ITS EXPRESSIBILITY.
CC - SUBUNIT: COMPONENT OF THE CHROMATIN-ASSOCIATED POLYCOMB COMPLEX
CC (PCG).
CC - SUBCELLULAR LOCATION: Nuclear.
CC - SIMILARITY: CONTAINS 1 A.T HOOK DNA-BINDING REPEAT.
CC - SIMILARITY: CONTAINS 1 CHROMO DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X62537; CAA44398.1; -
CC DR EMBL: Y13274; CAA73723.1; -
CC DR PIR: S23796; S23796.
CC DR HSSP: P2197; IAP0.
CC DR MGD: MGI:88289; Cbx2.
CC DR InterPro: IPR000637; AT_hook.
CC DR InterPro: IPR000953; Chromo.
CC DR Pfam: PF00385; chromo_1.
CC DR Pfam: PF02178; AT_hook_1.
CC DR SMART: SM00384; AT_hook; 1.
CC DR SMART: SM00298; CHROMO_1; 1.
CC DR PROSITE: PS00598; CHROMO_1; 1.
CC DR PROSITE: PS50013; CHROMO_2; 1.
CC DR Chromatin regulator; Nuclear protein; Transcription regulation;
KW Repressor; DNA-binding.
KW
FT DOMAIN 12 70 CHROMO.
FT DNA_BIND 75 87 A.T HOOK.
FT DOMAIN 102 120 SER-RICH.
FT DOMAIN 121 130 ASP/GLU-RICH (ACIDIC)
FT DOMAIN 164 169 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT CONFLICT 1 1 M -> F (IN REF. 2).
FT CONFLICT 85 85 A -> V (IN REF. 2).
FT CONFLICT 107 107 S -> G (IN REF. 2).
FT CONFLICT 133 133 R -> G (IN REF. 2).
FT CONFLICT 133 133
FT SEQUENCE 519 AA: 54890 MW: 761055F44423A80E CRC64;

```

Alignment Scores: 0.123 Length: 519  
Pred. No.: 120.50 Matches: 93  
Score: 36.468 Conservative: 51  
Percent Similarity: 23.548 Mismatches: 161  
Best Local Similarity: 2.688 Indels: 90  
Query Match: 1 Gaps: 15



## Alignment Scores:

Pred. No.: 0.359 Length: 982  
 Score: 115.50 Matches: 83  
 Percent Similarity: 38.258 Conservative: 57  
 Best Local Similarity: 22.688 Mismatches: 146  
 Query Match: 2.57% Indels: 80  
 DB: 1 Gaps: 16

US-09-806-276a-4 (1-2546) x HBL1\_CAEEL (1-982)

```

OY 227 AAGCTCTAGTGGCCGTAACCTCAGAGAGTCTCTTTGACCCCATTCATTTGAGACC 286
    |||||
DB 174 LysProGluSerThrSerValGluGly-----ThSerSerAsnGlnVal 189

OY 287 ACTTGTGAGCCACTGAGAGAAATGATGTGACAGTCTCTATCAAAAGATTCAGAAAC 346
    |||||
DB 190 ThrSerGluProVal----- 194

OY 347 ATATACCATCTGTGAGAAAGTGCCCTTTCCTCCGCTT--GCAAAATGACATTCACA 403
    :|||
DB 195 -----GlnMetProGlnMetProIleProValIleProSerPheLeu 208

OY 404 AATTCAAAATGCGCAGCAGACCCCAATTTAC---CTGAAGAGCCCAATTAACAAGAA 460
    :|||
DB 209 LysAsnSerLeuProIleProIleThrProThrGlnSerAlaAsnValGluArg 228

OY 461 GGAAGAAATTTAATCTGAGGAGCATCTGTCTCTGATATGATCATGCTCCCGCTTG-- 518
    :|||
DB 229 SerAsnSerProSerIleGluGluAlaLeu-LeuLeuThrLeuSerGluGlnPheAl 248

OY 519 -GAGACTTTCGCAACACCATTCACATTTGGCAAGAGGCCAGACATGCTTTGAGAT 577
    |||||
DB 248 agluValPheAlaGluAlaLysIleArgLysSerSer-----GluSe 264

OY 578 ATTTCCTTTCTTCAAGGAGGAGCTTCTTCTCTGTAAGAACAGAGAAAGACACACTG 637
    |||||
DB 264 rtleGlyIleGlnArgSerGlyThrSerAlaPheLeuAlaIleGluProLysGluMetSe 284

OY 638 GGGCAGTTCCTGGGGCATAATGAGTCTTCCGGGCCACAGCACCCTCGACTGTGTTTC 697
    |||||
DB 284 rmetSerSerAlaAsnAsnAsnAsnGluGluAlaProIleAlaSerThrValSerAlaGlySe 304

OY 698 ACAGAA-----CGCCCTCCCGGCTCAAAAATGACC 730
    |||||
DB 304 rThrProThrThrThrThrSerAlaSerPheCysArgPro-ProGlyLeuGlyProValA 324

OY 731 ATCTCCCTCCGACCATTTGAGAGA--TCCCAAGCTCTCATGTTGCCCTTATTG----- 781
    |||||
DB 324 lalaPro-ProThrGlnAsnGlyGlnThrProMetLeuValCysProIleCysGlyPhe 343

OY 782 -----TCACCAGTGACATTTAATTCACAAACAGAGATCCCTTGGGCCAGCAAGCTG 832
    |||||
DB 344 MetCysProSerLysPheHisPheAsnSerHisMetAsnThrHisGlyAspHisGlnCys 363

OY 833 CCCAGGCTTACTCGAGCCCGTCATGAGAGAAAACCTCAGAGAAAGACAGTCTGTTG 892
    |||||
DB 364 Ser-----MetCysAspTyrThrSerArgThrCysGluArgLeuLysLysHisMetArg 381

OY 893 GAGATGAGGACAGTCCAC-----CAGGAGACACTCGTGGGCC 931
    |||||
DB 382 GluSerHisThrValGluGluGlnLeuArgAlaGlyPheGluSerGluProAlaLysGlu 401

OY 932 TCCAGCGGTTTGCATCTCAGTCCAGCCAGAGAGACAGCAGCAGCTCCAGCCTGTC 991
    |||||
DB 402 SerAlaSerSerProLysAsnLeuSerLeuSerLysAspGlySerAlaThrSer----- 419

OY 992 GAACAGTACCCGACTGCCAGCCAGAGCATGTTTGACCATCCACCCCATGCGACCTC 1051
    |||||
DB 420 -----ProIleAsnGluThrPheAsnLeuSerThrThrMetAlaSer 433

OY 1052 ATCAGAGGAAGACTAAG-----TCAGAGAGATGCCCTCTGACCTTACAGGTTCC 1102
    |||||
DB 434 lIleuAspSerThrThrAsnAsnAlaValaSerThrThrThrGluGlnProSerAla 453
  
```

```

OY 1103 CTCCTCCCTGACAGCTTGATCTTTGG-----CCCTCACTTTTGATGAGC----- 1150
    |||||
DB 454 SerSerAlaLeuThrLeuAspMetSerSerThrProSerLeuLeuSerThrLeuAlaHis 473

OY 1151 -----CTGATGTATGATGATTAATAATAG-----TAACAGATGCCACATT 1192
    |||||
DB 474 SerSerPheGlyValSerAlaLeuAspGlnIleLysAlaIleSerGluAsnProSerPhe 493

OY 1193 TTTCTTTGGG 1204
    |||||
DB 494 MetProGluGly 497
  
```

## RESULT 9

```

ID N153_HUMAN STANDARD; PRT; 1475 AA.
AC P49790;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nuclear pore complex protein Nup153 (Nucleoporin Nup153) (153 kDa
DE nucleoporin).
GN NUP153.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=94154002; PubMed=8110839;
RA McMorris J., Bastos R., Horton H., Burke B.;
RT "Sequence analysis of a cDNA encoding a human nuclear pore complex
RT protein, hnp153."
RL Biochim. Biophys. Acta 1217:219-223(1994).
CC -!- FUNCTION: POSSIBLE DNA-BINDING SUBUNIT OF THE NUCLEAR PORE
CC COMPLEX (NPC). THE REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN
CC ANCHORING COMPONENTS OF THE PORE COMPLEX TO THE PORE MEMBRANE.
CC -!- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX. LOCATED TO THE
CC TERMINAL RING STRUCTURE OF THE NUCLEOPLASMIC CAGE.
CC -!- DOMAIN: CONTAINS F-X-F-G REPEATS.
CC -!- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
CC SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
CC NUP1, NSP1, POM 121 AND MAMMALIAN P62.
CC -!- SIMILARITY: CONTAINS 4 RANBP2-TYPE ZINC FINGERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL, Z25535; CAAB0982.1; -
DR Genbank, HGNC:8062; NUP153.
DR MIM: 603948; -
DR InterPro: IPR001876; Znf_RanBP.
DR Pfam: PF00641; zf_RanBP; 4.
DR SMART: SM00547; Znf_RBP2; 4.
DR PROSITE: PS01358; ZF_RANBP2_1; 4.
DR PROSITE: PS01359; ZF_RANBP2_2; 4.
KW Nuclear protein; transport; Repeat; Zinc-finger; DNA-binding.
FT DOMAIN 4
FT 447 GLY-RICH.
FT ZN_FING 657 687 POLY-GLY.
FT ZN_FING 722 751 RANBP2-TYPE 1.
FT ZN_FING 793 822 RANBP2-TYPE 2.
FT ZN_FING 851 880 RANBP2-TYPE 3.
FT ZN_FING 880 880 RANBP2-TYPE 4.
SQ SEQUENCE 1475 AA; 153889 MW; 3CB415A6909DF80E CRC64;
  
```

## Alignment Scores:

Pred. No.: 0.438 Length: 1475  
 Score: 115.00 Matches: 88

Percent Similarity: 36.58%  
 Best Local Similarity: 23.16%  
 Query Match: 2.56%  
 DB: 1  
 Conservative: 51  
 Mismatches: 141  
 Indels: 100  
 Gaps: 19  
 US-09-806-276a-4 (1-2546) x N153\_HUMAN (1-1475)

122 ATTCGTCGCGCCGCGGAGCGCCGCGGAGCGC----- 154  
 814 ValSerCysMetSerGluLeuPheProAlaSerSerSerThVal 833  
 155 ---GCTGGGATCCTCAGCGCGCGCGGCTTCTCTGCTGTCGTCAGACTGATGAT 211  
 834 ProValSerLeuProSerGlyGlySerLeuGlyLeuGlyLysPheLysProGluGly 853  
 212 AACTG-----CTCTAGAGAGCGCTC-----TTGGCC 241  
 854 SerTrpAspCysGluLeuCysLeuValGlnAsnLysAlaAspSerThrLysCysLeuAla 873  
 242 GTAACTCAGAGAGGTTCTCTTTGAGCCCATCTCATTTGAGAGCC-----ACTTGTGAA 295  
 874 CysGluSerAlaLysProGly---ThLysSerGlyPheLysGlyPheAspThrSerSer 892  
 296 GCCACTTGAGAAATATGATGAGAGTTCCATCAAAAAGGATTGAGAAACATATACCAT 355  
 893 SerSerSerAsnSerAlaAlaSerSerSerPheLysPheGly----- 906  
 356 CTGTGAAGAAAGTGGCCCTTCTCCCGCTTGCAAAATAGACATTCGCAATTCGCAATG 415  
 907 -----ValSerSerSerSerGly 913  
 416 CCAGCCAGAGCCCAATTATCTGAAAGAGCCAAATACAGAAAGAGAAAGAA----- 469  
 914 ProSerGlnThr-----LeuThrSerThrGlyAsnPheLysPheGlyAspGlnGly 931  
 470 TTAAAGTGAAGGACATTCTGTCCTCCGATATGATGATGATGATGATGATGATGATGAT 511  
 932 PheLysIleGlyValSerSerAspSerGlySerLeuAsnProMetSerGluGlyPheLys 951  
 512 -----CCGCTTGAGACTTTCGCGCACACCATC-----CAC 541  
 952 PheSerLysProIleGlyAspPheLysPheGlyValSerSerLysLysProGluGlu 971  
 542 ATGGCAGAGAGGCGGAGCAGCAT-----GTCTTGAGAGATATTCCTTCTTCAAGG 595  
 972 ValLysLysAspSerLysAsnAspAsnPheLysPheGlyLeuSerSerGlyLeuSerAsn 991  
 596 AACTAGAGCTTTACTCTGGAACCAAGAGAGAGACACCTGCGGAGTTCCTCGGCGCAT 655  
 992 ProValSerLeuThrProPheGlnPheGlyValSerAsnLeuGlnGluLysLys 1011  
 656 AATGAGTTCTCCGCGGAGCAGACCTCGGACTCT-----GTGTTCAAGAA 703  
 1012 GluGluLeuProLysSerSerSerAlaGlyPheSerPheGlyThrGlyValIleAsnSer 1031  
 704 AGCCCTCCCGGCTG-----CTAAAATGCCATTCCTCCCGC 742  
 1032 ThrProAlaProAlaAsnThrIleValThrSerGluAsnLysSerPheAsnLeuGly 1051  
 743 ACCATTGAGAGATCCCAAGCTCATGTTGCTTATGTCACAGTACATTTATATTC 802  
 1052 ThrIleGluThrLysSerAlaSerValAlaProPheThrCysLysThrSerLysAlaLys 1071  
 803 AAACAGAG-----TCCTTCGG-----CCAGCAAG 829  
 1072 LysGluGluMetProAlaThrLysGlyLysPheSerPheGlyAsnValGluProAlaSer 1091  
 830 CTGCCAGGCTTACCTGAGAGCCGCTC-----ATGAGAGAGAAAGCTCAGAG----- 877  
 1092 LeuProSerAlaSerValPheValLeuGlyAlaGlyThrGluGluLysGlnGluProVal 1111  
 878 AAAAGCAGCTCTTGAGAGATGAGACAGTCCACAGGAGACACC----- 922

Db 1112 ThrSerThrSerLeuValPheGlyLysLysAlaAspAsnGluGluProLysCysGlnPro 1131  
 QY 923 -----TCGTGGGCGCTCCAGCGGTCTGCATCTCAGTCCAGCAAGGAGAGACACCAC 976  
 Db 1132 ValPheSerPheGlyAsnSerGluGlnThrLysAspGluAsnSerSerSerThrPhe 1151  
 QY 977 TCTCCAGCGCTG-----TCCGAACAGTACCCGAGCTGCGCAGCGAGCAGCATGTT 1027  
 Db 1152 SerPheSerMetThrLysProSerGluLysGluSerGluGlnProAlaLysAlaThrPhe 1171

RESULT 10  
 ZEP2\_HUMAN  
 ID ZEP2\_HUMAN STANDARD: PRT; 1833 AA.  
 AC P31629;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Human immunodeficiency virus type I enhancer-binding protein 2  
 GN HIV-EP2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91217105; PubMed=2022670;  
 RA Nomura N., Zhao M.-J., Nagase T., Maekawa T., Ishizaki R., Tabata S.,  
 RA Ishii S.;  
 RT "HIV-EP2, a new member of the gene family encoding the human  
 RT immunodeficiency virus type 1 enhancer-binding protein. Comparison  
 RT with HIV-EP1/EPDII-BP1/MBP-1." J. Biol. Chem. 266:8590-8594(1991).  
 RL [2]  
 RP SEQUENCE OF 1184-1323 FROM N.A.  
 RX MEDLINE=91062349; PubMed=2247438;  
 RA Rustgi A.K., Van'T Veer L.J., Bernards R.;  
 RT "Two genes encode factors with NF-kappa B- and H2TF1-like DNA-binding  
 RT properties." Proc. Natl. Acad. Sci. U.S.A. 87:8707-8710(1990).  
 RL [3]  
 RP FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE  
 CC 5'-GGGAGTTC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF  
 CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.  
 CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS  
 CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I  
 CC MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT  
 CC IN T-CELL ACTIVATION.  
 CC SUBCELLULAR LOCATION: Nuclear.  
 CC INDUCTION: BY MITOGEN AND PHORBOL ESTER.  
 CC SIMILARITY: STRONG, TO HIVEP1.  
 CC

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC

EMBL: M60119; AB88218.1; -  
 DR EMBL: M61744; AA36202.1; -  
 DR PIR: A39829; WMDHE2.  
 DR HSSP: P15822; IERO.  
 DR TRANSFAC: T00939; -  
 DR Genew: HGNC:4921; HIVEP2.  
 DR MIM: 143054; -  
 DR InterPro: IPR000822; Znf\_C2H2.  
 DR Pfam: PF00096; Zf-C2H2; 2.  
 DR SMART: SM00355; Znf\_C2H2; 2.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 2.  
 DR PROSITE: PS00157; ZINC\_FINGER\_C2H2\_2; 2.  
 DR Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;  
 KW Nuclear protein; Repeat.

FT DOMAIN 324 330 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT 369 369 SER-RICH.  
 FT DOMAIN 1286 1310 ASP/Glu-RICH (ACIDIC).  
 FT ZN\_FING 1186 1208 C2H2-TYPE.  
 FT ZN\_FING 1214 1238 C2H2-TYPE.  
 SQ SEQUENCE 1833 AA; 202128 MW; BAF0FA37BE075C23 CRC64;

## Alignment Scores:

Prod. No.: 0.464 Length: 1833  
 Score: 115.00 Matches: 82  
 Percent Similarity: 36.88% Conservative: 36  
 Best Local Similarity: 25.62% Mismatches: 96  
 Query Match: 2.56% Indels: 106  
 DB: 1 Gaps: 14

US-09-806-276a-4 (1-2546) x ZEP2\_HUMAN (1-1833)

QY 334 AGGATTTCAGAAACATATACCTCTGTGAAGAAAGTGGCCCTT----- 375  
 DB 724 ArgileglnThrHisValProSerTyrglySerValMetTyrrHisSerIleSerGlnIle 743  
 QY 376 -----TCGCCGCT-----TGCAAAATGACATCTCAATTCGCAAAAT 414  
 DB 744 LeuGlyGlnHisSerProAlaIleValIleCysIlyValAsp-----Gluasn 759  
 QY 415 GCCAGCCAGAACCCCAATTTACCTGAAGCAGCCCAATTAACAGAAAGAAAGAAATTTAA 474  
 DB 760 MetThrGln-ArgThrLeuValThrAsnAlaIle-----MetGlnGlyIleGlyPheas 777  
 QY 475 ACAGAGGAGACATCTCTCTCCATATATGATCACTCCCGCGCTGGAGACTTGGCCACAC 534  
 DB 777 nileAlaGlnValLeu----- 782  
 QY 535 CATCCACATTTGGCAAGAGGCGCAGACGATGCTTTGGAGATATTTCTTTCTTCAAG 594  
 DB 783 -----GlyGlnHis-AlaGlyLeuGlnIlyTyrrProIleTyrIlysa 796  
 QY 595 GAACATAGACCTTTTACTTCTGGAAGAACGAGAAAGACACCTGGCGACCTTCCCTGGCA 654  
 DB 796 laProGlnThrLeu-----ProleuGlyL 804  
 QY 655 TAATGATCTCTTCGCGG-----CCAAAGACACCTGAGCTGTGTTCACAGAAAC 705  
 DB 804 engIlySerIleProleuCysLeuPro-SerThrSerAspSerVal----- 819  
 QY 706 GCCCTCCCGGTGCTCAAAATGCAATCTCCCTCCGACCATGTGGAGATCCCAAGCTGC 765  
 DB 820 -----AlaThrLeuGlyGlySerIlySerIlySargme 828  
 QY 766 CATGTGGCCCTTATTGTACACGAGACATTT-----AATTCCAAGAGAGATCTTCCGGGCC 822  
 DB 828 tLeuSerProAlaSerSerIleuGlnLeuPheMetGlnThrIlyGlnIlyS-----Ar 846  
 QY 823 AGCAAGCTGCCAGGCTTAGCTGCGAGCCGCTCATGAGGAAAGAAAGCTCAGAG----- 877  
 DB 846 gValIlyGlnIlyuysmetTyrglyGlnIleValGlnIleuSerAlaValaGlnIlyu 866  
 QY 878 -----AAAGACAGCTGTGGAGATG 900  
 DB 866 rAsnSerAspIleIlySAspIleuSerArgProGlnIlySProGlnIlyValaGlnIly 866  
 QY 901 GACATCCACAG-----GGAGACACCTCGTGGGCTCCAGCGGTTCGTGATCTCA 951  
 DB 886 yCysAlaSerGlnProIlySAspGlyLeuGlnSerGlySerSerSerPheSerSerIle 906  
 QY 952 GTCCAGCCAGCAGACAGACACCTCTCCAGCTTCCGCAAGATACCCCGACCTGGCC 1011  
 DB 906 rProSerSerGlnAspTyrrProSerValSerProSerSerArgIlyProPheProPr 926  
 QY 1012 AGCCGAGCAGCATGTTGAC-----CATCCACCCAGCAGCAGCATCATCAAGGAAA 1062  
 DB 926 oSerIlySgluMetLeuSerGlySerArgAlaProleuProIlyGlnIlySerserGlyPr 946

QY 1063 GACTACTCAGAGAGATCC-----CTCTGACCTTAC 1095  
 DB 946 oSerGlnSerIlySgluSerSerAspGlyLeuAspIleAspGlnThrAlaSerAspIle 966  
 QY 1096 AGGTTCCCTCCCTCCCTGCAGCTTGTGGAGCCCTCACTTTGGATGAG 1147  
 DB 966 rMetSerProGlnSerSerSerIleuProAlaGlyAspGlyGlnIlyGlnIly 983

RESULT 11  
 ID YAC SCHPO STANDARD; PRT: 596 AA.  
 AC 009729;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C31A2.12 in chromosome 1.  
 GN SPAC31A2.12.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 NX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymopre B.,  
 RA Welfens I., Vanstreels E., Rieger M., Schefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goftau A., Cadieu E., Dreano S., Gloux S., Laure V., Motier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Usery D., Batteil B.G., Nurse P.,  
 RT "The genome sequence of Schizosaccharomyces pombe."  
 RL Nature 415:871-880(2002).  
 CC -1- SIMILARITY: TO S.POMBE SPAC8A4.13C AND TO YEAST ROD1, YER022W AND  
 YGR068C.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isdb.ch/announce/>  
 CC or send an email to [license@isdb.ch](mailto:license@isdb.ch)).

CC EMBL: Z50113; CAA90470.1;  
 CC DR Hypothetical protein; Transmembrane.  
 KW TRANSMEM 194 211 POTENTIAL.  
 FT DOMAIN 532 537 POLY-SER.  
 SQ SEQUENCE 596 AA; 65902 MW; C93B9FA55688E7B1 CRC64;

Alignment Scores:  
 Prod. No.: 0.65 Length: 596  
 Score: 111.50 Matches: 97  
 Percent Similarity: 33.57% Conservative: 44



Best Local Similarity: 23.10% Mismatches: 139  
 Query Match: 2.48% Gaps: 140  
 DB: 1 Caps: 20

US-09-806-276a-4 (1-2546) x YAMC\_SCHPO (1-596)

```

QY 234 ACTTGGCCGTAACCTGAGAA-----GGT 257
DB 179 ThrtPrpAsnLysIleGluYrThrIleSerValProThrLysAlaIleIle 198
QY 258 TCTCTTGGACCC--CATCTCATTGAGGCACTTGTGAG-- 296
DB 199 SerTyrIleProIleHisValLeuValProLeuLeuYsArgLeuThrIleGly 218
QY 296 ----- 296
DB 219 IleSerIleThrLeuYsGluTyrIleThrIleHisValAlaHisGlyTyrAsnGlyLeu 238
QY 297 CCACCTGAGAAAATGATGTC-----ACAGTTCCTATCAAAAG 335
DB 239 ProAlaSerLysAspLeuValArgThrValArgSerLeuGlnThrGluLeuGlu 258
QY 336 GATTGAGAAACATATACCATCTGTGAGAAAGTGGCCCTTCTCCGCT 384
DB 259 PheSerAspHisTyrGluLeu-ThrLysAsnLeuLeuProSerSerLeuValGly 278
QY 385 -----TGCAAATATGACATTTCTCAATTCCAAAATGCGCCAGACCCCATTT 434
DB 278 sleuGlnAspCysAspLeuAspGlyIleLysIleArgHis-LysLeuLysPheSerVal 298
QY 435 ACCTGAAGACGACCCAAATACAGAAAGAAAGAAATTTAAACGTGAGGACATTCTGTC 494
DB 298 erLeuArg-----AsnProAspGlyHisIleSerGluLeuArgAlaIleLeuPro 315
QY 495 CTGATATCATAGTCCCGCTTGAGACTTTCCGACACCATCATCATGGCAAGAG 554
DB 315 alValLeuMetIlePro----- 321
QY 555 GCCAGCAGATGCTTTGGAGAT-----ATTTCCTTTCTTCAAGGAACTACAGAC 605
DB 322 -----GlnLeuPheGlyAspArgValGluValGluSerLeuArgGluAsnPheGlu 339
QY 606 TT-----TTTCTGGAACCCAGAGAGAAAGACACACCTGGGCCAGTCCCTGGGC 653
DB 339 erPheAsnGlnProLeuProSerTyrGlnAsnSerMetLysArgLeuTyrAsp-Gly 358
QY 654 ATATGAGTCT-----TCCGGCCAGACAGCACCCTGG----- 686
DB 359 LeuSerTyrSerAsnLeuAspThrProLeuProSerGlyAlaThrThrProArgArg 378
QY 686 ----- 686
DB 379 AspSerMetGlnProThrTyrAlaSerAsnGlnAlaHisArgArgGlnLeuIleAlaGly 398
QY 687 -----ACTCTGTGTCACAGAAACCCCTCC--CGTGCTCAAAATAGCATCTCC 737
DB 399 LeuThrGlnLeuAlaLeuGlnGlnProGlnArgSerProAsnGlnHisSerPro 418
QY 738 TCCCGACCATTTGGAGAT-----CCC-----AAGCTCTCATGTTGCCCTATTGTAC 785
DB 419 SerAsnHisProGlnAspPheProProSerPheSerLeuGlySerAsnProSerGly 438
QY 786 CAGTGACATTTAATTCACACAGAGTCTTCCGGCCAGCAAGCTGCCAGCTTAC 845
DB 439 AlaAlaSerAlaValIleThrArgAsnProSer--GluThrSerLeuAlaAspLeuSer 457
QY 846 GCGAGCCCGTCATGAGAGAAAAGCTCAGAGAAAAGAGCTGTGGAGAAATGGCAG 905
DB 458 ArgValProSerTyrLysGluAlaThrArgSerAlaValProLeuGlnSerProLeuGln 477
QY 906 TCCACGAGGAGACACCTCGTGGGCTCCAGCGTTCTGCATCTCAGT--CCAGCCAG 962
DB 478 SerThrLeuProSerTyrGlnAspValAlaArgGlnLeuHisLeuSerArgProProSer 497

```

```

QY 963 GCAGAGACACCACTCTCTCA-----GCCGTGCCAGAC-----AGTACCCGACGCGC 1010
DB 498 ProGlyIleValThrProProGlnArgThrSerProSerPhePheValSerProThrIle 517
QY 1011 CAGCCGAGACATGTTG----- 1028
DB 518 SerThrArgGlnSerLeuAspSerArgArgSerPheGlnHisSerThrSerSerSer 537
QY 1029 ---ACATCCGACCCCATCGAGCTCATCAAGGAAAGAACTAAGTACAGAGAGT---CCC 1082
DB 538 GlyIleSerProSerHisSerSerAlaSerAlaHisLeuSerGlnAlaSerAsnPro 557
QY 1083 TCTCTGACCTTCAGGTTCTCCCTCCCTGCACCTTGATCTGGCCCTCAG 1136
DB 558 AsnGlySerSerSerAlaProHisArgProThrSerAlaArgValGlySerHis 575

RESULT 12
ID KDGL_HUMAN STANDARD: PRT: 1117 AA.
AC Q13574; O00542;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Diacylglycerol kinase, zeta (EC 2.7.1.107) (Diglyceride kinase) (DGK-zeta) (DAG kinase zeta).
GN DGKZ OR DAGK6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE=Endothelial cells;
RX MEDLINE=96215319; PubMed=8626588;
RA Bunting M., Tang W., Zimmerman G.A., McIntyre T.M., Prescott S.M.;
RT "Molecular cloning and characterization of a novel human
RL J. Biol. Chem. 271:10230-10236(1996).
RN [2]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RC TISSUE=Skeletal muscle;
RX MEDLINE=97303161; PubMed=9159104;
RA Ding L., Bunting M., Topham M.K., McIntyre T.M., Zimmerman G.A.,
RA Prescott S.M.;
RT "Alternative splicing of the human diacylglycerol kinase zeta gene in
RL muscle.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:5519-5524(1997).
RN [3]
RP PHOSPHORYLATION, AND SUBCELLULAR LOCATION.
RX MEDLINE=98379993; PubMed=9716136;
RA Topham M.K., Bunting M., Zimmerman G.A., McIntyre T.M.,
RA Blackhear P.J., Prescott S.M.;
RT "Protein kinase C regulates the nuclear localization of diacylglycerol
RL kinase-zeta.";
RL Nature 394:697-700(1998).
RN [4]
RP FUNCTION: DISPLAYS A STRONG PREFERENCE FOR 1,2-DIACYLGLYCEROLS
OVER 1,3-DIACYLGLYCEROLS, BUT LACKS SUBSTRATE SPECIFICITY AMONG
MOLECULAR SPECIES OF LONG CHAIN DIACYLGLYCEROLS.
CC CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol = ADP + 1,2-
diacylglycerol 3-phosphate.
CC SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC TISSUE SPECIFICITY: HIGHEST LEVELS IN BRAIN, AND SUBSTANTIAL
LEVELS IN SKELETAL MUSCLE, HEART, AND PANCREAS.
CC PTM: PHOSPHORYLATION OF THE MARKS HOMOLOG DOMAIN BY PKC REDUCES
NUCLEAR ACCUMULATION OF DGK-ZETA.
CC SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE
FAMILY.
CC SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHOSBO-ESTER AND DAG
BINDING DOMAINS.
CC SIMILARITY: CONTAINS 2 ANK REPEATS.

```



[illegible]

```

OY 537 TCCACATTTGGCAAGAGGGCCAGCAGATGTCCTTGGAGATATTTCTTCTTCAAGGA 596
Db 131 -----
OY 597 ACTACGAGCTTTTACCTGGAAACAGAGAAACACACCTGGCCAGTTCCTGGGCATA 656
Db 132 -----
OY 657 ATGAGTTCTTCGGGCCACAGACACCTCGACTGTGTTCACAGAAACCCCTCCCGG 716
Db 144 ProlyAspValAlaPro-----ProProPro 152
OY 717 TGTCAAAAATGCCATCTCCCTCCGACATTGGAGATCCAGCTCATGTTCGCTT 776
Db 153 ValSerLys---ProSerGluProArgProSerProGluProGlnLeuSerLys----- 169
OY 777 TATGTCCACAGTACATTAAATTCGAAACAGAGAGCTTCGGGCCAGCAAGCTGCCA 836
Db 170 -----
OY 837 GGC-----TTAGCTCGAGCCCGCTCATGTGAGGAAAAAGCTCAGGAAAAAGCAGTC 887
Db 178 GlyThrLeuArgPheArgLeuSerProAlaAlaArgAsnIleLeuGlnLysHisSerLeu 197
OY 888 TGTTCGAGAAATGGACAGCTCCAGGAGGAGACACCTGCTGG----- 929
Db 198 -----AspAlaSerGlnGlyThrAlaThrGlyProArgGlyIlePheThrLysGluAsp 215
OY 930 -----GCTCCAGCGGTCTGTCGATCCAGTCACATCCAGCCAG 962
Db 216 AlaLeuLysLeuValGlnLeuLysGlnThrGlyLysIleThrGluSerArgProThrPro 235
OY 963 GCAGAGACAGACCCACTCTCCAGCC----- 986
Db 236 AlaProThrAlaThrProThrAlaProSerProLeuGlnAlaThrAlaGlyProSerTyr 255
OY 987 -----TGTCCGAAACGTACCCCGACTGGCCAGCCAGG-----ACA 1022
Db 256 ProArgProValIleProProValSerThrPro---GlyGlnProAsnAlaValAlaGlyThr 274
OY 1023 TGTTCACATCCACCCATCCAGCTCATCAACAGGAAAGAGATGACAGAGATGCC 1082
Db 275 PheThrGluIlePro-----AlaSerAsnIleArgValIleAlaLysArgLeuThr 292
OY 1083 TCTCTGACCTTACAGGTTCCCTCTCTCTCTCTCTGACCTTGATTTGGCCCTTCTGG 1142
Db 293 GluSerLysSerThrVal-----ProHisAlaTyr 302
OY 1143 ATGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1202
Db 303 AlaThr-AlaAspLysAsp----- 308
OY 1203 GGTAAAGGTACAAAACAACTAACACAGTTGAAGAGAGAGGCTTCGAGCTGTATT 1262
Db 309 -----LeuGlnAlaValLeu 313
OY 1263 TGCAGTTTGTGTGGTTTCTAAATATATCTTACAACTATTTTCTTACCTGTTA 1322
Db 313 uLysValArgGlnAspLeuValLysAspPhe-----LysValSerValAsnAspH 331
OY 1323 TGCCCTGTTCGCAAAAACAAATTTGAAAAAACAACAAGCA-----AACCT 1370
Db 331 eLleLysAlaIleAlaVal-----ThrLeuLysGlnMetProAspValAsnVa 348
OY 1371 ATCTTGGCAAAAAAGAGAGTGAAGTCAAGCCCATTT-----TCAAGAGGCAATTGG 1421
Db 348 lSerTrpAspGlyGlyGlyProLysGlnLeuProPheLysAspLysSerValAlaValAla 368
OY 1422 TGAATGTCG-----CTCACATATTTGTTTGCAGACACACAGAAATCTGGCTGGCAG 1475
Db 368 aThrAspLysGlyLeuLeuThrProIleLysAspAlaAlaLysGlyIleGlnG 388
OY 1476 GATTGGCACTAGCTATGAAGGCTGAGGAGTCAACATTAAGAACTTCAGAGAACTTAT 1535

```

```

Db 388 uLleAlaAspSerValLysAlaLeuSerLys----- 398
OY 1536 AGCACTCCGACATTTCTGAGCAGAGAGAGAGTCAAAATTTATTTACACCTTAAGCCTTTT 1595
Db 399 -----LysAlaArgAspGlyLysLeuLeu----- 406
OY 1596 TGTAGACTCTTTCTATATATTTGCTTAGGCTCACCATTCGCAATTCCTCAGGTAAAC 1655
Db 407 -----ProGluLysGlnGlyLys 414
OY 1656 TTTTCTGTTTTCATTTGACATTTATGAGTTTGGGATTTCTTCTGATCTTATATA 1715
Db 414 rPheSerIleSerAsnLeuGlyMetPheGlyLysAspGlnPheThrAlaValIleAsnPr 434
OY 1716 TCCCTATATATTATNCTA-----TATTCGAAAAATTTGACTGTCCAGC----- 1758
Db 434 oProGlnAlaCysIleLeuAlaValAlaGlyArgPheArgProValLeuLysLeuThrGluAs 454
OY 1759 -----TACATGTTGTTAGACACAGCAGCAAGATTTACTGTAAT 1797
Db 454 pGluGluGlyAsnAlaLysLeuGlnIleArgGlnLeuIleThrValThr 470

RESULT 14
V7OK_TYMVA STANDARD: PRT: 628 AA.
AC P20131:
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE 69 kDa protein.
OC Turnip yellow mosaic virus (Australian isolate).
OC Viruses; ssRNA positive-strand viruses, no DNA stage: Tymovirus.
OX NCBI_TaxID=12155;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=90021184; PubMed=2800335;
RA Reese P., Mackenzie A., of the genome of an Australian isolate of turnip
RT Nucleotide sequence of the genome of an Australian isolate of turnip
RL Virology 172:536-546(1989).
CC -1- FUNCTION: NOT KNOWN.
CC -1- SIMILARITY: TO 65 to 70 kDa PROTEIN FROM OTHER TYMOVIRUSES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: J04373; AAA46591.1;
CC DR PIR: J00110; J00110; TYMO_45_70kDa.
CC DR InterPro: IPR004935; TYMO_45kDa;
CC Pfam: PF03251; TYMO_45kDa; 70kDa;
CC SEQ ID: 628 AA; 68740 MW; 67CD342A09161D96 CRC64;

Alignment Scores:
Pred. No.: 0 864 Length: 628
Score: 110.00 Matches: 85
Percent Similarity: 30.87% Conservative: 133
Best Local Similarity: 22.43% Mismatches: 129
Query Match: 2.44% Indels: 19
Gaps: 19

US-09-806-276a-4 (1-2546) x V7OK_TYMVA (1-628)
OY 230 CCTCACTTGCCCGCTACCTCAGAGAGTTC-----TCTTGACCCCATCTCATTTTC 280
Db 197 ProHisSerLeuArgProGlyArgSerArgSerArgGlnLeuGlnProThr-IleArg 216
OY 281 GAAGCCACTCTTGAAGCCACTTGAAGAAATGAT-----GTGACAGTTCTCTATCAAAA 334

```



```

RT amphipathic structure as determined by NMR spectroscopy."
RL Biochemistry 34:3022-3031(1995).
[9]
RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE-91221459; PubMed-1902642;
RA Gitschler J.;
RT "The molecular basis of hemophilia A.";
RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
[10]
RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE-89088506; PubMed-2491949;
RA White G.C. II, Shoemaker C.B.;
RT "Factor VIII gene and hemophilia A.";
RL Blood 73:1-12(1989).
[11]
RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
RX MEDLINE-95245332; PubMed-7728145;
RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
RL Hum. Mutat. 5:1-22(1995).
[12]
RP VARIANT GLN-2326.
RX MEDLINE-86235434; PubMed-3012775;
RT "Identification of a missense mutation in the factor VIII gene of a mild hemophiliac.";
RL Science 232:1415-1416(1986).
[13]
RP VARIANT PRO-2135.
RX MEDLINE-88096539; PubMed-3122181;
RA Levinson B., Janco R.L., Phillips J.A. III, Gitschler J.;
RT "A novel missense mutation in the factor VIII gene identified by analysis of amplified hemophilia DNA sequences.";
RL Nucleic Acids Res. 15:9797-9805(1987).
[14]
RP VARIANT GLN-2228.
RX MEDLINE-88191889; PubMed-2833855;
RA Youssoufian H., Antonarakis S.E., Bell W., Griffin A.M., Kazazian H.H.;
RT "Nonsense and missense mutations in hemophilia A: estimate of the relative mutation rate at CG dinucleotides.";
RL Am. J. Hum. Genet. 42:718-725(1988).
[15]
RP VARIANT GLY-291.
RX MEDLINE-88220354; PubMed-2835904;
RA Youssoufian H., Wong C., Aronis S., Platokoukis H., Kazazian H.H. Jr., Antonarakis S.E.;
RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution in exon 7 of the factor VIII gene.";
RL Am. J. Hum. Genet. 42:867-871(1988).
[16]
RP VARIANT CYS-1708.
RX MEDLINE-89274393; PubMed-2499363;
RA O'Brien D.P., Tuddenham E.G.;
RT "Purification and characterization of factor VIII 1,689-Cys: a nonfunctional cofactor occurring in a patient with severe hemophilia A.";
RL Blood 73:2117-2122(1989).
[17]
RP VARIANT CYS-391.
RX MEDLINE-90001543; PubMed-2506948;
RA Shima M., Ware J., Yoshioke A., Fukui H., Fulcher C.A.;
RT "An arginine to cysteine amino acid substitution at a critical thrombin cleavage site in a dysfunctional factor VIII molecule.";
RL Blood 74:1612-1617(1989).
[18]
RP VARIANT LEU-189.
RX MEDLINE-90057680; PubMed-2510835;
RA Chan V., Chan T.K., Tong T.M., Todd D.;
RT "A novel missense mutation in exon 4 of the factor VIII:C gene resulting in moderately severe hemophilia A.";
RL Blood 74:2688-2691(1989).
[19]
RP VARIANT LEU-2326.
RX MEDLINE-89197216; PubMed-2495245;
RA Inada H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
RT "Mild hemophilia A resulting from Arg-to-Leu substitution in exon 26 of the factor VIII gene.";
RL Hum. Genet. 81:335-338(1989).
[20]
RP VARIANT HIS-391.
RX MEDLINE-89264602; PubMed-2498882;
RA Arai M., Inada H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr., Fujimaki M., Hoyer L.W.;
RT "Direct characterization of factor VIII in plasma: detection of a mutation altering a thrombin cleavage site (arginine-372-->histidine).";
RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
[21]
RP VARIANT CYS-1708.
RX MEDLINE-90105723; PubMed-2104766;
RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr., Phillips J.A. III, Janco R.L., Hoyer L.W.;
RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to Cys) in the factor VIII gene of two unrelated patients with cross-reacting material-positive hemophilia A.";
RL Blood 75:384-389(1990).
[22]
RP VARIANTS GLN-2228 AND LEU-2326.
RX MEDLINE-90123183; PubMed-2105106;
RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G., Mancuso G., Morfini M., de Biasi R., Baudo F., Carbonara A.;
RT "Recurrent mutations and three novel rearrangements in the factor VIII gene of hemophilia A patients of Italian descent.";
RL Blood 75:662-670(1990).
[23]
RP VARIANT CYS-391.
RX MEDLINE-90329422; PubMed-1973901;
RA Pattinson J.K., McVey J.H., Boon M., Afari A., Tuddenham E.G.;
RT "CRM+ haemophilia A due to a missense mutation (372-->Cys) at the internal heavy chain thrombin cleavage site.";
RL Br. J. Haematol. 75:73-77(1990).
[24]
RP VARIANTS PHE-1699 AND CYS-1708.
RX MEDLINE-90152691; PubMed-2105906;
RA Higuchi M., Wong C., Kochan L., Olek K., Aronis S., Kasper C.K., Kazazian H.H., Antonarakis S.E.;
RT "Characterization of mutations in the factor VIII gene by direct sequencing of amplified genomic DNA.";
RL Genomics 6:65-71(1990).
[25]
RP VARIANTS CYS-1728 AND ASP-1941.
RX MEDLINE-90169988; PubMed-2106480;
RA Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E., Kazazian H.H.;
RT "Use of denaturing gradient gel electrophoresis to detect point mutations in the factor VIII gene.";

```

## Alignment Scores:

Pred. No.:	1 22	Length:	2351
Score:	110.00	Matches:	79
Percent Similarity:	38.51%	Conservative:	55
Best Local Similarity:	22.70%	Mismatches:	122
Query Match:	2.44%	Indels:	92
DB:	1	Gaps:	17

US-09-806-276a-4 (1-2546) x FA8\_HUMAN (1-2351)

OY	285	CCACTTCGAGACCACTGAGAAAATGATGTCACAGTCTTCATCAAAAAGCATTCAGANA	344
DB	768	ProSerThrArgInLysGlnPheAsnAlaTrnThrIlePro-GluAsnAspIleGluLys	787
OY	345	A-----CATATACCATCTGTGAAGAAAGTGGCCCTTTC	377
DB	787	sthrAspProTrpPheAlaHisArgThrPrometProLysIleGlnAsnValSerSerSe	807

```
QY 378 TCCCGCTTGCAAAATAGACATTCATTCCTCAAAATGCCAGCCCAAGCCCAATTAC 437
Db 807 r-----AspleuMetLeuArgInser-ProThrProHisGly 822
QY 438 TGAAGCAGCCCAATTAACAGAAAGAAATTTAACTGAGGAGCATTCCTCTG 497
Db 822 euserLeuSerAspLeuIngluAlaLysTyrGluThrPheSerAspAspProSerProG 842
QY 498 ATATGATCAGT-----CCCCCGCTGGAGACTTTCCGCACACCATTCCACA 542
Db 842 lylAlaIleAspSerAsnAsnSerLeuSerGluMetThrHisPheArgProGluLeuHis- 861
QY 543 TTGGCAAGAGAGCCACAGCATGCTCTTGGAGATATTCCTT----- 586
Db 862 -----Hisser-----GlyAspMetValPheThrProGluSerGly 874
QY 587 -----CTTCAAGGAACTACGACTTTTACCTGGAACAGGAGAAAGACACCTGGGC 641
Db 874 eugInLeuArgLeuAsnGluLysLeu-----GlyThrThrAlaAlaThrGluLeuLysL 892
QY 642 AGTTCCCTGGCATATGAGTTCTTCCGGGCCACAGCAGCTCGAGCTGTCTTCACAG 701
Db 892 yslLeuAsp-----PheLysValSerSerThrSerAsnAsnLeuLysert 907
QY 702 AAACGCCCTCCCGGTGCTC-----AAAAATGCCATCTCCCTCCGACCATTG 749
Db 907 hrlLeProSerAspAsnLeuAlaIaGlyThrAspAsnThrSerSerLeuGly----- 924
QY 750 GAGGATCCCAAGCTCTCATGTTGCCCTTATGTACACAGTACATTAATCCAAACAG 809
Db 925 -----ProProSerMetProValHisTyrAspSerGluLeuA 937
QY 810 ACTCC-----TTGGGGCCAGCAAGCTGCCAGGCTT--ACCTGGAGCCCGTCATGG 860
Db 937 sprThrLeuPheGlyLysLysSerSerProLeuThrGluSerGlyGlyProLeuSerL 957
QY 861 AGGAAAAAGCTCAGGAGAAAGCAAGTCTGTGGAGATGGAGACAGTCCACGAGGAGACA 920
Db 957 euserGluGluAsnAsnAspSerLysLeuLeuGluSerGlyLeuMetAsnSerGluGlu 977
QY 921 CCTGTGGGGCTCCAGCGGTTCTGCATCTCAGTCCAGC-----CAAGCAGAGACA 971
Db 977 erserTrpGlyLysAsnValSerSerThrGluSerGlyArgLeuPheLysGlyLysArg 997
QY 972 GCCACTCTCCAGCTGCTCCGAACAGTACCCGACTGGCCAGCCGAGGACATGTTGACC 1031
Db 997 laHisGlyProAlaLeuLeu-----ThrLysAspAsnAlaLeuP 1010
QY 1032 ATCCCAACCCCATGCCAGCTCATCAAGGAGAAAGACTAAGTCAGAGAGTCCCTCTGAC- 1090
Db 1010 helysValSerLieserLeuLeuLysThrAsnLysThrSerAsnAsnSerAlaThrAsnA 1030
QY 1091 -----CTTACAGGTTCCCTCCCTCCCTCCCTGACGCTGATCTTGGGCCCTCACTT 1139
Db 1030 rGlySerThrHisLaspolyPro-----SerLeuLeuIleGluAsnSerProSerValT 1048
QY 1140 TGGATGAGGTGCTGAAT----- 1156
Db 1048 rpgInAsnIleLeuGluSerAspThrGluPheLysValThrProLeuIleHisAspA 1068
QY 1157 -----GTAATGATATAAAAT 1171
Db 1068 rgmMetLeuMetAspLysAsn 1074
```

Search completed: July 18, 2003, 17:27:00  
Job time : 75.4741 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 18, 2003, 17:08:12 ; Search time 180.354 Seconds

(without alignments)  
5817.411 Million cell updates/sec

Title: US-09-806-276a-4

Perfect score: 4500

Sequence: 1 cggggctagccggagacc.....cagctanaagagggcagta 2546

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-O/cgn2.1/USPFCPOOL/US09806276/runat\_18072003\_160924\_12753/app.query.fasta.1.3854  
-DB=SPREMBL\_21 -QEXT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR.SCORE=DCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09806276.ecgn.1.1.397@runat.18072003\_160924\_12753 -NCPU=6 -ICPU=3  
-NO MAP -LARGEOUTERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: SPREMBL\_21:\*  
2: SP.archaea:\*  
3: SP.bacteria:\*  
4: SP.fungi:\*  
5: SP.human:\*  
6: SP.invertebrate:\*  
7: SP.mammal:\*  
8: SP.mhc:\*  
9: SP.organelle:\*  
10: SP.phage:\*  
11: SP.plant:\*  
12: SP.potent:\*  
13: SP.virus:\*  
14: SP.vertebrate:\*  
15: SP.unclassified:\*  
16: SP.virus:\*  
17: SP.bacteriap:\*  
18: SP.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1326	29.5	254	4	Q9UK12

2	1320	29.3	254	4	Q95353	Q95353 homo sapien
3	1319	29.3	254	4	Q9UDJ0	Q9UDJ0 homo sapien
4	1230	27.3	254	11	Q9COC5	Q9COC5 mus muscultu
5	407.5	9.1	210	4	Q14613	Q14613 homo sapien
6	406.5	9.0	210	4	Q9UN50	Q9UN50 homo sapien
7	406.5	9.0	210	4	Q9BS13	Q9BS13 homo sapien
8	216	4.8	409	11	Q9DBM1	Q9DBM1 mus muscultu
9	216	4.8	409	11	Q91W92	Q91W92 mus muscultu
10	159	3.5	605	12	Q9O5L3	Q9O5L3 ceratophth
11	150	3.3	326	10	Q22514	Q22514 santalum al
12	143	3.2	1044	4	Q9P206	Q9P206 homo sapien
13	137	3.0	580	4	Q9UEF3	Q9UEF3 homo sapien
14	135	3.0	2766	11	Q9OZ88	Q9OZ88 mus muscultu
15	132.5	2.9	791	4	Q9BY92	Q9BY92 mus muscultu
16	131	2.9	148	4	Q8TB51	Q8TB51 mus muscultu
17	130.5	2.9	741	11	Q89015	Q89015 mus muscultu
18	130.5	2.9	1337	5	Q960Z6	Q960Z6 drosophila
19	130.5	2.9	1373	5	Q917F7	Q917F7 drosophila
20	129	2.9	673	11	Q91VK6	Q91VK6 mus muscultu
21	126	2.8	1188	10	Q41805	Q41805 zea mays (m
22	124	2.8	150	11	Q9OZT9	Q9OZT9 mus muscultu
23	124	2.8	307	4	Q9NT06	Q9NT06 mus muscultu
24	124	2.8	319	4	Q9UE55	Q9UE55 mus muscultu
25	123	2.7	406	2	Q8VPM0	Q8VPM0 mus muscultu
26	123	2.7	741	11	Q8R080	Q8R080 mus muscultu
27	122.5	2.7	150	11	Q9Z0X0	Q9Z0X0 mus muscultu
28	122.5	2.7	520	11	Q35731	Q35731 mus muscultu
29	122	2.7	474	5	Q8T1B6	Q8T1B6 dictyostell
30	121.5	2.7	543	4	Q14879	Q14879 mus muscultu
31	121	2.7	351	10	Q39492	Q39492 chlamydomon
32	121	2.7	628	12	Q56253	Q56253 turnip yell
33	120	2.7	551	3	Q14269	Q14269 schizosach
34	120	2.7	1315	10	Q9SPM0	Q9SPM0 zea mays (m
35	119	2.6	1006	11	Q62501	Q62501 rattus norv
36	117.5	2.6	242	4	Q9BRP6	Q9BRP6 mus muscultu
37	117	2.6	501	4	Q96FV8	Q96FV8 mus muscultu
38	116.5	2.6	1200	5	Q9VFH4	Q9VFH4 drosophila
39	116.5	2.6	1413	5	Q9N9M9	Q9N9M9 leishmania
40	116	2.6	370	10	Q9FN39	Q9FN39 arbidopsis
41	116	2.6	437	12	Q69146	Q69146 human hepe
42	116	2.6	1212	4	Q8TAT0	Q8TAT0 mus muscultu
43	115.5	2.6	544	11	Q8VCG2	Q8VCG2 mus muscultu
44	115	2.6	616	4	Q9H6K5	Q9H6K5 mus muscultu
45	115	2.6	984	4	Q8TDE4	Q8TDE4 mus muscultu

## ALIGNMENTS

RESULT 1  
Q9UK12 PRELIMINARY: PRT: 254 AA.  
AC Q9UK12  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE CRIB-containing BORG2 protein (Hypothetical 27.7 kDa protein).  
GN BORG2 OR DKFZP434A0530.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=HEART;  
RX MEDLINE=99421943; Pubmed=10490598;  
RT Jobery G., Perlungher R.R., Macara I.G.;  
RT "The Borgs, a new family of Cdc42 and Tc10 GTPase-interacting  
RT proteins.";  
RL Mol. Cell. Biol. 19:6585-6597(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RA Blum H., Bauersachs S., Mewes H.W., Gassenhuber J., Wiemann S.;

RN	Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RP	[3]
RC	SEQUENCE FROM N.A.
RA	Strasbourg R.;
RL	submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AF164118; AAD4815.1; -
DR	EMBL; AL136842; CAB6776.1; -
DR	EMBL; BC019270; AAH19270.1; -
DR	InterPro: IPR000095; PAKbox/Ribondng.
DR	SMART: SM00285; PBD: 1.
KW	Hypothetical protein.
SQ	SEQUENCE 254 AA; 27678 MW; C3ICDBEL5IC16IED CRC64;
 Alignment Scores:	
Pred. No.:	2,19e-118
Score:	1326.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	29.47%
DB:	Gaps: 0
 US-09-806-276A-4 (1-2546) x Q9UKI2 (1-254)	
OY	413 ATGCCAGCCAAAGACCCCATTATTCCTGGAAGACCCTTAACAAGAAAGAATAATT 472
Dy	1 MetProAlalyslrProiletyrLeuylsalalaasnaenlyslgilyllysPhe 20
OY	473 AAATGGAGGGCAATTCGTGTCGCATATGATCATAGTCCCCGGCTTGAGACTTTGGCCAC 532
Dy	21 LysleuAlayaplleleuserProaspheIlleserPropleGlyaspPheargHis 40
OY	533 ACATCCACATTTGGCAAAGAGGGCCAGCAGCATGTCTTTGGAGATATTTCTCTTCAA 592
Dy	41 ThrillestlllegilylsGluglglmlhisaspValPheglyaspIlleserPheleuIn 60
OY	593 GGGAACTAACGAGCTTTTAACCTCGAAACACAGAGAAACACACACCTGGCCAGTCCCTGG 652
Dy	61 GlysnnyrGlualeuleuProglYasnlnGlysalahlsleuGlyglnPheProGly 80
OY	653 CATATGAGTTCCTCCGGGCCCAACAGCACCTCGGACTCTGTGTTACAGAAAGCCCTCC 712
Dy	81 HisnngluPhephearglalaInsenserThrserspserValPheThrGluThrProser 100
OY	713 CCGGTGCTCAAAAAATGCCATGTCCCTCCGACCATGTGAGAGATCCCAAGCTCTCATGTG 772
Dy	101 ProvalleuylsaanaIalleSerleuProThrlllegilylserGlnlaIleuwlleu 120
OY	773 CCTTATTTGTACACAGTGACATTTAATTCCAAACAGAGAGTCTTCGGGCCAGAAAAGTG 832
Dy	121 ProteuleuserProvalThrPheasnserlysglnGlnuserPheglyProAlalysleu 140
OY	833 CCCAGGCTTAGCTGCGAGCCCGTCATGGAGGAAAAAGSTCAGGAGAAAAGCATGTCTGTG 892
Dy	141 ProArgleuserCyglurProvalMetIuglnylsalaglnglnlylserSerleu 160
OY	893 GAGATGGGAAGTCCACCCAGGGAGAACCCCTGGGGGCTCCAGGGGTTCGATCTCAG 952
Dy	161 GlusnnglynPhalnIselnGlysprhsseTrtpglyserGelyselAlaserIn 180
OY	953 TCCAGCCAAAGCAAGACAGCCACTCTCCAGCTGTCCGAACAGTACCSCGACTGGCCA 1012
Dy	181 SersegrnglnyldgaPsrserhlsserserserleusergluglnlylProasPrpro 200
OY	1013 GCCGAGACATGTTTGACATTCSSACCCCATGCGACTCATCAAGGAAAGACTAAGTCA 1072
Dy	201 AlaglusPmetrPheasPhIsProthrProcsglnleuIllelysgilylThrlyser 220
OY	1073 GAGAGTCCCTCTCGACTTAAAGGTTCCCTCCCTCCCGGAGCTTGATCTTGAGGCC 1132
Dy	221 GluglnuserleuserAspleuThelGylserleuuserleuGlnleuaspleuGlyPro 240
OY	1133 TCACTTTTGGATGAGTCTCAATGTATGATGAATAAATAAG 1174

Db	241	SerLeuLeuAspValLeuAsnValMetAspLysAsnLys	254
RESULT 2			
095353	PRELIMINARY:	PRT;	254 AA.
AC	095353;		
DT	01-MAY-1999 (TrEMBLrel. 10, Created)		
DT	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	MS55-related protein.		
GN	UBL.		
OC	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BRAIN;		
RA	MEDLINE=98204843; PubMed=9535835;		
RA	Alberts A.S., Bouquin N., Johnston L.H., Treisman R.;		
RT	"Analysis of Rho-binding proteins reveals an interaction domain		
RT	responsible in heterotrimeric G protein beta subunits and the yeast		
RL	J. Biol. Chem. 273:8616-8622(1998).		
DR	J. Biol. Chem. 273:8616-8622(1998).		
DR	InterPro: IPR000095; PAKbox/Rhobndng.		
DR	SMART: SM00285; PBD: 1.		
SEQ	SEQUENCE 254 AA; 27738 NM; C31CDBE15IDA605D CnC64;		
Alignment Scores:			
Pred. No.:	8.28e-118	Length:	254
Score:	1320.00	Matches:	253
Percent Similarity:	99.61%	Conservative:	0
Best Local Similarity:	99.61%	Mismatches:	1
Query Match:	29.33%	Indels:	0
DB:	4	Gaps:	0
US-09-806-276A-4 (1-2546) x 095353 (1-254).			
QY	413	ATGCCAGCCCAAGACCCCAATTTACSTGAAAGCGCCCAATTAACAAGAAAGAAATTT	472
Db	1	MetProAlaLysThrProIleuLeuLysAlaAlaAsnAsnLysGlyLysPhe	20
QY	473	AAACTGAGGACACTTGTGTCTCTGATATGATCAATCCCCCGCTTGAGACTTCCGCAC	532
Db	21	LysLeuAlaTrpAspLysLeuSerProAspMetLysSerProLysGlyAspPheAlaGln	40
QY	533	ACCATCCACATTTGGCAAAAGAGGCGACACGATGCTTTGGAGATTAATTTCCCTTCCAA	592
Db	41	ThrIleHisIleLysGlnGlyGlnHisAspValPheGlyAspLysPheLeuGln	60
QY	593	GGGAATAAGAGCTTTTACCTGTGAACACGAGAAAGACACACTGGGGCAGTTCCSTGG	652
Db	61	GlyAsnTrpGlyLeuLeuProCysGlnGlnCysAlaHisLysLeuGlyGlnPheProGly	80
QY	653	CATAATGAGTTTCCGGGCGCAAGACACACTGGACTGTGGTACAGAAAGCCCTCC	712
Db	81	HisAsnLeuPhePheArgLysAsnSerThrSerAspSerValPheThrGlyThrProSer	100
QY	713	CCGGTGTCAAAATCCATCTCCCTCCCGACATTTGAGAGANTCCCAAGCTCTCATGTTG	772
Db	101	ProValLeuLysMetLysLeuLeuLeuLeuProThrIleGlyGlyLysSerGlnAlaLeuMetLeu	120
QY	773	CCCTTATGTGCACACTGATGATTAATTTCAAACGAGAGTCTTGGGCGACAAAGCTG	832
Db	121	ProLeuLeuSerProValThrPheAsnSerLysGlnCysLeuSerPheGlyProAlaLysLeu	140
QY	833	CCGAGCGTTAGTGGCGAGCGCTGCATGGAGGAAAGAGTCGAGGAAAGACAGCTGTG	892
Db	141	ProArgLeuSerCysLeuProValMetGlnGlnLysAlaGlnGlnLysSerSerLeuLeu	160
QY	893	GAGAAATGGACACTCCACACGAGGAGACACACTTGTCGGGCTCCACAGGTTTGCACTTCAG	952

D <sub>b</sub>	161	GLUASNGLYThValHisIsoleuLysPheThrProGlySerSergylSerAlaSerGln	180
Q <sub>y</sub>	953	TTCAGCGCAGGACGAGACAGCCACTCTCCAGCCTGTCCACAGTACCCTGGATGCCA	1017
D <sub>b</sub>	181	SerSerGlnGlyArgAspSerHisSerSerSerLeuSerGlnGlnTyrProAspTrpPro	200
Q <sub>y</sub>	1013	GCCGAGGACATGTTTGACCATCCACACCCCATGCGAGCTCATCAAGGAAAGCAATAAGCA	1077
D <sub>b</sub>	201	AlaGluAspMetPheAspHisProThrProCysGlnLeuLeuLeuGlySerIleThrIleSer	220
Q <sub>y</sub>	1073	GAGGAGTCCTCTCTGCACCTTACAGGTTCCCTCCCTCCCTGGAGAGCTGATTCTTGAGCCC	1132
D <sub>b</sub>	221	GluGlnSerLeuSerAspLeuThrGlySerLeuLeuPheLeuGlnIleuAspLeuLeuPro	240
Q <sub>y</sub>	1133	TCACCTTTGGATGAGTGCTGAATGATTAATGATTAATAAATAAG	1174
D <sub>b</sub>	241	SerLeuLeuAspGluValLeuAsnValMetAspLysAsnLys	254

RESULT 3	Q9U0J0	PRELIMINARY:	PRT:	254 AA.
ID	Q9U0J0			
AC	Q9U0J0:			
DT	01-MAY-2000	(TREMBLrel. 13,	Created)	
DT	01-MAY-2000	(TREMBLrel. 13,	Last sequence update)	
DT	01-MAR-2002	(TREMBLrel. 20,	Last annotation update)	
OS	Cdc42 effector protein 3.			
OC	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-PREGNANT UTERUS.			
RX	MEDLINE=21125609; PubMed=11035016;			
RA	Hirsch D.S., Plitene D.M., Burbelo P.D.;			
RT	"A new family of Cdc42 Effector Proteins, CEPs, Function in Fibroblast			
RL	and Epithelial Cell Shape Changes.";			
RN	J. Biol. Chem. 276:875-883(2001).			
RP	[2]			
RC	SEQUENCE FROM N.A.			
RC	TISSUE-PREGNANT UTERUS;			
RA	Thomas D.S., Burbelo P.D.;			
RT	"Unique domains define a new family of Cdc42 effector proteins			
RL	(CEPs).";			
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF104857; AAD16888.1; .			
DR	InterPro: IPR000095; PAKbox/RhoIndng.			
DR	SMART: SM00285; PBD: 1.			
SO	SEQUENCE 254 AA: 27679 MW: 29236600 Da			

Alignment Scores:	
Pred. No.:	1,03e-117
Score:	131e+00
Percent Similarity:	99.61%
Best Local Similarity:	99.61%
Query Match:	29.31%
DB:	
Length:	254
Matches:	253
Conservative:	0
Mismatches:	1
Indels:	0
Gaps:	0

QY 413 ATGCCAGCCAGACCCCAATTATCTGGAAGCAGCCAAATAACAAGAAAGAAAGAAATTT 472  
 |||||  
 Db 1 MetProAlaIalThrProIleTyrLeuLysAlaIaAsnAsnLysLysLysPhe 20  
 |||||  
 QY 473 AAAC TAGGAGACATTCCTGCTCTCATATMGATCAGTCCCGCGCTTGAGACTTTGGCCAC 532  
 |||||  
 Db 21 LysLeuArgAspIleuLeuSerProAspMetIleSerProProLeuGlyAspPheArgHis 40  
 |||||  
 QY 533 ACCATCCAAATTCGCAAAAGAGGGCCAGCAGCATGCTTTGAGATATATTCCTTTCCAA 592  
 |||||  
 Db 41 ThrIleHisIleGlyLysGluGlyGlnHisAspPheIleGlyLysPheAsnSerPheLeuGln 60

QY	593	GGGACATACAGACCTTTTAACTGGAAACAGAAAGACACCTGGGACCTGCTGGG	552
Db	61	GLYASNTRYGLULEULEUPROGLYASNGINGLUYALAHNLSLEUGLGNPHEPTGUY	80
QY	653	CATPATAGAGTTCTTCCGGGCCAACAGACCTGGAGCTGTGTTCACAGAAACCCCTCC	712
Db	81	HISASNGIUPHERHEATGALAAANSETHIRSEASPSEVALPHEHTGUTHPROSER	100
QY	713	CCGGGCTCAAAAAATGCCATCTCCCTCCGACCATTTGGAGATGCCAAGCTTCATGTTG	772
Db	101	PROVALLEUYASNAIALIASELLEUPROTHRIEGLUYSERGINALALEULEU	120
QY	773	CCCTATTGTGCACAGAGCATTTAATGCCAAGCAGAGTCTTCCGGGCCAGCAAAAGCTG	832
Db	121	PROLEULEUSERPROVALTHRIHEANSERLUSGNGIUSERPHEGLYPROALATLUSLEU	140
QY	833	CCGAGGCTTGGCAGCCACGCCCTCATGAGAGAAAAAGCTCAGAGAAACACATCTGTGTTG	892
Db	141	PROATRGLEUSERYSGIUPROVALMETGLUGIUSYALAGIINGLUYSERSEULEU	160
QY	893	GAGAAATGGACAGCTCCACCAGGAGAACACCTCGGGGGGCCAGCGGTTCTGCATCTCAG	952
Db	161	GLUASMDIYTHRYVALHISGLINGLUYSPHRHETTRPGLYSERSEGLYSERLAASERIN	180
QY	953	TTCAGGCCAAAGGAGAGACAGACGCACTCTCCAGCCTGTGCCAGACAGTACCACATGGGCA	1012
Db	181	SETHSERFINGIYARGASPSERHISSETHSERSETHSERLEUSERGLUGIINTYPRKASPIRPRO	200
QY	1013	GCCGAGACATGTTTGACCATCCACACCCATCGAGGTCATCAAGGAAAGACTAAGTCA	1072
Db	201	ALAGLUASPMETRHEASPHISPROTHRPROCYSGIULEULEUYSEGLUYSTHLYSESER	220
QY	1073	GAGAGATCCCTCTGCAGACTTACAGGTTCCCTCTCTCCCTGACACTTGATCTTGGGCC	1132
Db	221	GLUGIUSERLEUSERASPHEUTHRGUYSETHLEUSERLEUGIHLLEUASPHEUGLYPRO	240
QY	1133	TCACITTTTGATGAGAGCTGTAATGTATAGTAAATAAATAAG	1174
Db	241	SETHLEUASPPGLUVALLEUASVALMETASPLASIVASIVUS	324

RESULT 4	
09C0C5	PRELIMINARY; PRT; 254 AA.
ID 09C0C5	
AC 09C0C5	
DT 01-JUN-2001	(TREMBLrel, 17, Created)
DT 01-JUN-2001	(TREMBLrel, 17, Last sequence update)
DT 01-MAR-2002	(TREMBLrel, 20, Last annotation update)
DE 3200001F04Rik	protein (RIKEN CDNA 3200001F04 gene).
GN 3200001F04Rik	
OS Mus musculus (Mouse)	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
NCBI_TaxID=10090;	
[1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD, AND EMBryo.	
RX MEDLINE=21085660; Pubmed=11217851;	
RA Kawai J., Shingawa A., Shibata K.,	
RA Arakawa T., Hara A., Fukunishi Y., Yoshino M., Itoh M., Ishii Y.,	
RA Atzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,	
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,	
RA Kodota K., Watsuda H.A., Ashburner M., Batalov S., Casavant T.,	
RA Frieschmann W., Gaasterland T., Gissi C., Ring B., Kochiya H.,	
RA Kuehl P., Lewis S., Matsuo Y., Nikidol J., Pesole G., Quackenbush J.,	
RA Schrieml L.M., Stubi F., Suzuki R., Tomita M., Wagner L., Washio T.,	
RA Sakai K., Okido T., Fununo M., Aono H., Baldarelli R., Bash G.,	
RA Blake J., Boffelli D., Bojunga N., Caranci P., de Bonaldo M.F.,	
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,	
RA Gastincich S., Hill D., Hornam M., Hume D.A., Kamiya M., Lee N.H.,	
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,	
RA Moridone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,	
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,	
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmberg L.,	

100





```

Db      91 ArgValGlyValProProArgArg-----MetalSerProAlaAlaProSer 106
QY      692 GTGTTACAGAAAGCGCTCCCGGTCGTCACAAATGCGCATCTCCCTCCGACATTGGA 751
Db      107 ProAlaProProProIleSerProIleIleIleIleIleIleIleIleIleIleIle 126
QY      752 GGATCC-----CAAGCTCTCATGTTGCCCTTATGTCA-----CCAGTGACA 793
Db      127 GluAlaThrTyrAspSerLeuValMetGlyLysLeuSerPheAspSerThrProAlaSer 146
QY      794 TTTAATTCACAAAGAGAGTCCTTCGGG-----CCAGCAAGCTG 832
Db      147 SerThrAspGlyHisSerGlyTyrGlyLeuGluSerGlyPheCysThrIleSerArgLeu 166
QY      833 CCCAGGCTTAGCTGCGAGCCCGCATGAGAGAAAGATCAGAGAAAGACAGTCGTG 892
Db      167 ProArgVal-----GluLysHis----- 172
QY      893 GAGAAATGGGACAGTCACACAGGAGACACCTGCTGGGCTCCAGCGGTTCTGCATCTCAG 952
Db      173 -----SerAsn 174
QY      953 TCCAGCCAAAGCAGACAGACACCACTCTCCAGCGCTCCGAAACAGTACCCGACGTGGCA 1012
Db      175 ArgAspArgAspArgAspProAspHisSerGlnAspArgGluGln---SerSerPhePro 193
QY      1013 GCCGAGAGCATGTTTGACCATCCACCCCATGGAGCTCATCAAGGAGAAAGACTAAGTCA 1072
Db      194 SerGlu-----ProThrPro-----AsnPro 200
QY      1073 GAGAGATCCCTCTGACCTTACAGGTTCCCTCCCTCCCTGAG-----CTTGATCTT 1126
Db      201 GluLeuArgArgSerAsp-----SerLeuLeuSerPheArgPheAspLeu 217
QY      1127 GGGCCCTCACTTTTGATGATGAGTGTGATGATG 1162
Db      218 GlyProSerLeuLeuSerGluLeuGlyValMet 229

RESULT 9
QY      091W92 PRELIMINARY; PRT; 409 AA.
AC      091W92;
DT      01-DEC-2001 (TREMBLrel. 19, Created)
DT      01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      RIKEN cDNA 1810058K22 gene.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN      NCBI_TaxID=10090;
RP      [1]
RC      SEQUENCE FROM N.A.
RA      TISSUE=SALIVARY GLAND;
RA      Strausberg R.;
RL      Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC016250; AAH16250.1;
SQ      SEQUENCE 409 AA; 43096 MW; 1794FDD0850171F0E CRC64;

Alignment Scores:
Pred. No.: 1.18e-11 Length: 409
Score: 216.00 Matches: 82
Percent Similarity: 42.86% Conservative: 26
Best Local Similarity: 32.54% Mismatches: 62
Query Match: 4.80% Indels: 82
DB: 11 Gaps: 14

US-09-806-276a-4 (1-2546) x 091W92 (1-409)
QY      488 CTGCTCCTGATATGATCCCGCTGGAGACTTGCACACATCATTCATGGC 547
Db      33 LeuthrAlaAspMetIleSerProProIleGlyAspPheArgHisThrMetHisValGly 52
QY      548 AAAGAGGGCCAGCAGATGCTTTGGAGATATTTCTTTCTT----- 589

```

```

Db      53 ArgGlyGly-----AspValPheGlyAspThrSerPheLeuSerAsnHisGlyArg 70
QY      590 CAAGGAACTCAGAGGCTTTTACCTGGAAC-----CAGGAAAGCA 631
Db      71 SerGlyAsnThrHisArgSerProArgSerPheLeuAlaArgLysLeuGlnValArg 90
QY      632 CACCTGGCCGAGTTCCTCCGCGCATAAAGATTCCTCCGGCCAAACAGACACCTCGACTCT 691
Db      91 ArgValGlyValProProArgArg-----MetalSerProAlaAlaProSer 106
QY      692 GTTTCACAGAAAGCGCTCCCGGTCGTCACAAATGCGCATCTCCCTCCGACATTGGA 751
Db      107 ProAlaProProProIleSerProIleIleIleIleIleIleIleIleIleIleIle 126
QY      752 GGATCC-----CAAGCTCTCATGTTGCCCTTATGTCA-----CCAGTGACA 793
Db      127 GluAlaThrTyrAspSerLeuValMetGlyLysLeuSerPheAspSerThrProAlaSer 146
QY      794 TTTAATTCACAAAGAGAGTCCTTCGGG-----CCAGCAAGCTG 832
Db      147 SerThrAspGlyHisSerGlyTyrGlyLeuGluSerGlyPheCysThrIleSerArgLeu 166
QY      833 CCCAGGCTTAGCTGCGAGCCCGCATGAGAGAAAGATCAGGAGAAAGACAGTCGTG 892
Db      167 ProArgVal-----GluLysHis----- 172
QY      893 GAGAAATGGGACAGTCACACAGGAGACACCTGCTGGGCTCCAGCGGTTCTGCATCTCAG 952
Db      173 -----SerAsn 174
QY      953 TCCAGCCAAAGCAGACAGACACCACTCTCCAGCGCTCCGAAACAGTACCCGACGTGGCA 1012
Db      175 ArgAspArgAspArgAspProAspHisSerGlnAspArgGluGln---SerSerPhePro 193
QY      1013 GCCGAGGACATGTTTGACCAATCCACCCCATGGAGCTCATCAAGGAGAAAGACTAAGTCA 1072
Db      194 SerGlu-----ProThrPro-----AsnPro 200
QY      1073 GAGAGATCCCTCTGACCTTACAGGTTCCCTCCCTCCCTGAG-----CTTGATCTT 1126
Db      201 GluLeuArgArgSerAsp-----SerLeuLeuSerPheArgPheAspLeu 217
QY      1127 GGGCCCTCACTTTTGATGATGAGTGTGATGATG 1162
Db      218 GlyProSerLeuLeuSerGluLeuGlyValMet 229

RESULT 10
QY      0905L3 PRELIMINARY; PRT; 605 AA.
AC      0905L3;
DT      01-MAY-2000 (TREMBLrel. 13, Created)
DT      01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE      01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE      EBNM2-like protein (EBNA-2).
OS      ceropitilicline herpesvirus 15.
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Gammaherpesvirinae; Lymphocryptovirus.
RN      NCBI_TaxID=104228;
RP      [1]
RC      SEQUENCE FROM N.A.
RA      MEDLINE=20057932; PubMed=10590127;
RA      Peng R., Gordanze A.V., Fuentes Panama E.M., Wang F., Zong J.,
RA      Hayward G.S., Tan J., Ling P.D.;
RT      "Sequence and functional analysis of EBNA-LP and EBNA2 proteins from
RT      nonhuman primate lymphocryptoviruses.";
RL      J. Virol. 74:379-389(2000).
RN      [2]
RC      SEQUENCE FROM N.A.
RA      STRAIN=LCL8664;
RA      MEDLINE=97048062; PubMed=8892903;
RA      Franken M., Devergne O., Rosenzweig M., Annis B., Kieff E., Wang F.;
RT      "Comparative analysis identifies conserved tumor necrosis factor

```

RT receptor-associated factor 3 binding sites in the human and simian  
RX Epstein-Barr virus oncogene LMP1.";  
RL J. Virol. 70:7819-7826(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LCL8664;  
RX MEDLINE=99412410; PubMed=10482645;  
RA Rivallier P., Quink C., Wang F.;  
RT "Strong selective pressure for evolution of an Epstein-Barr virus  
LMP2B homologue in the rhesus lymphocytovirus.";  
RL J. Virol. 73:867-8872(1999).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LCL8664;  
RX MEDLINE=20304984; PubMed=10846073;  
RA Jiang H., Cho Y.G., Wang F.;  
RT "Structural, functional, and genetic comparisons of Epstein-Barr virus  
nuclear antigen 3A, 3B, and 3C homologues encoded by the rhesus  
lymphocytovirus.";  
RL J. Virol. 74:5921-5932(2000).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LCL8664;  
RX MEDLINE=20440633; PubMed=10970361;  
RA Rao P., Jiang H., Wang F.;  
RT "Cloning of the rhesus lymphocytovirus viral capsid antigen and  
Epstein-Barr virus-encoded small RNA homologues and use in diagnosis  
of acute and persistent infections.";  
RL J. Clin. Microbiol. 38:3219-3225(2000).  
RN [6]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LCL8664;  
RX MEDLINE=21602573; PubMed=11739708;  
RA Rivallier P., Jiang H., Cho Y.-G., Quink C., Wang F.;  
RT "Complete Nucleotide Sequence of the Rhesus Lymphocytovirus: Genetic  
Validation for an Epstein-Barr Virus Animal Model.";  
RL J. Virol. 76:421-426(2002).  
RN [7]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LCL8664;  
RA Mognaddam A., Koch J., Annis B., Wang F.;  
RT Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
RL [8]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LCL8664;  
RA Mognaddam A., Annis B., Wang F.;  
RT Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
RL [9]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LCL8664;  
RA Rivallier P., Quink C., Wang F.;  
RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RL [10]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LCL8664;  
RA Rivallier P., Quink C., Wang F.;  
RT Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RL [11]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LCL8664;  
RA Rao P.V., Jiang H., Wang F.;  
RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RL [12]  
RP SEQUENCE FROM N.A.  
RC STRAIN-LCL8664;  
RA Rivallier P., Jiang H., Cho Y.-G., Quink C., Wang F.;  
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL: AF200187; AAF25206.1; -;  
DR EMBL: AY037858; AK95414.1; -;  
DR InterPro: IPR002965; P-rich extensn.  
DR InterPro: IPR000633; Vinculin\_2.  
DR PRINTS: PRO1217; PRICHTEXTENS.  
DR PRINTS: PRO0806; VINCULIN.

Seq	Sequence	605 AA:	64712 MW:	AD0F77111208AA4E CRC64:
Alignment Scores:				
Pred. No.:	3,98e-06	Length:	605	
Score:	159.00	Matches:	105	
Percent Similarity:	29.22%	Conservative:	44	
Best Local Similarity:	20.59%	Mismatches:	158	
Query Match:	3.53%	Indels:	204	
DB:	12	Gaps:	23	
US-09-806-276a-4 (1-2546) x Q905L3 (1-605)				
QY	18	CCCCGGCCACCGGCGCTGGGGCCCTTACGACCCCT-CTCGAGACGGATATATGCGGTGAGC	74	
DB	88	ProProValGlnProProProProProGlnAlaArgAlaGlyValGlyThrGlnLysProAspPro	10	
QY	75	AGGACACACACCGGCGGACACTGGCGTGGATCTGGCGACACGGGACGGATGCGTCGCGCC	13	
DB	108	ArgAsp-ArgAspPro	11	
QY	135	CGCGGAGCCCGGGGCGAGCGGTGGGATGCTCAGCGGCGCGCGGCTTGGTCGG-188		
DB	113	LeuSlyLysProGlyGlySerAlaValSerProLysAlaVal	12	
QY	189	-----TTGTGCTACAGACTGGATGATGACTGGCTCTA	22	
DB	126	ArgMetLeuCysMetSerAsnLeuIleValAlaArgGlnAlaIleGlyValAlaArgGlyLeuLeu	14	
QY	225	GGAAGCGCTCACTGGCGCGTAACTCGTAGGAAGGCTCTTGGACCCCATCTGATTTCGAAG	28	
DB	146	AlaThrLeuLys-----ProGlnAlaArgAlaGlyAsnProThrProProProMetAlaSer	16	
QY	285	CCACTT-CTGAAGCACTTGAAGAAATGATGTGACAGTCTCT-320		
DB	163	ProIleThrLeuSerProGlySerGlnAsnValSerProThrProProProSerThrPro	18	
QY	327	-----ATCAAAAGGATTCAGAAACATATACCATC-356		
DB	183	ProArgAlaAlaIleProSerGlyLeuGlnLeuProGlnGlyProGlnLysProGlnThrLeuAsnProGly	20	
QY	357	-----TGTGAAAGATGGGCGCTTCTCCGCGCTTGCACAAATATGACATTTCCAAAT	40	
DB	203	AsnHisThrAlaIleAspProGlnThrProThrMetProProProSerHisThrGlnSer	22	
QY	408	CCA-----AAATGCCAGCCA-422		
DB	223	ProThrAlaIleProCysArgProProProProGlnIleMetProProProSerAlaIleProThr	24	
QY	423	AGACCCCAATTTCTCTGAAGACGACCAATATACAGAAAGAAATTTAACTGAGGG	48	
DB	243	Arg-ProserProProArgProProGln-----251		
QY	483	ACATTCGTCTCTGATATGATCACTCCCGCGCTTGGAGACTTTCGACACCAT-537		
DB	252	-----ProHisThrAlaTyrAlaGlyProHisThrGlyThrGlnI	26	
QY	538	-----CCCATTTGG-566		
DB	262	HisThrGlnProProThrProProProTyrAlaArgProProProThrLysHisGlnAlaProThr	28	
QY	561	ACGATGCTTTGGAGATATTTCTCTTCAAGGACATAGACAGCTTTTACCTGGAAC	62	
DB	282	ArgValAlaProThrArgProSerProValGlnAlaIleProProAlaIle-----ProThrGlyThr	30	
QY	621	AGGA-----GAAGACACACTGGGCGCACTTCCTGGGCAAT-656		
DB	301	AspProThrLysGlnGlnAlaIleProProProAlaProThrProThrLysProSerProThrLysGln	32	
QY	659	GAGTCTCTCGGCGGACAGCACTCGCGGCTGTGGTTCACAGAAACGCTCCCGGATG	71	
DB	321	nglnAlaIleProProProAlaIleProProLeuThrSerProGlnHisProAspProGlnIle	34	
QY	719	CTCAAAATTCGATCTCC-737		



```

Db 341 rThProHisProHrProProLyHisIleProValHisProLeuThHisAlaProGl 361
QY 738 -----TC 739
Db 361 uGlySerProLeuGlnGlnAlaProGlnGlnLeuThProArGAsnLeuProProLeuSe 381
QY 740 CCGACCATTTGAGAGATCCCAAGCTCTCATGTTGCCCTTATTTGTCACAGTACATTAT 799
Db 381 rSerProLeuIlyValProThrValGlnLeuProProLeuGlyS-----ThrThle 398
QY 800 TCCAAACAGAGCTCTGGGGCCAGCAAGCTGCCAGGCTTAGCT-----845
Db 398 uProProProAlaProSer-----GluProAlaProGlyIleValArGThrGlnLeuGl 416
QY 846 -----GGAGCCGCTCA-----TGGAGAAAAAGCTCAGGAGAAAGCAGTCTG 889
Db 416 nProProAlaProProGlyProValArGThrProProValThrArGSerIlySAlaSerIly 436
QY 890 TTGAGACATGGAGAGTCCACAGGAGACA-----920
Db 436 sArGArGlySgIlySerArGAlaProSerArGAlaGlnIlySgIlySgIlySgIly 456
QY 920 -----920
Db 456 sGlyIlySAlaIlyAlaArGAlaArGLeuGlyGlnIlySerThProProGlnAlaGlyPr 476
QY 921 -----CCTCGTGGGCTCCAGCGTTTGCATCTCATGTCAGCCAGGC 964
Db 476 oValProSerProGlnProThrProHrProSerProSerMetProAlaLeuSerProAlaIlePr 496
QY 965 AGAGACAGCCACTCTCTCCAGCTCTGCGAAGTACCCGAGTCCAGGAGGAGGAGCAGTGTG 1024
Db 496 oLeuAlaSerArGnProValArGAlaProAsnSerSerProThrProGlyProSerGlnAl 516
QY 1025 TTTGACCATCCACCCCATGCGATCATCAAGGAGAAAGTCAAGTCAAGAGAGTCCCTC 1084
Db 516 aleu-----ProAlaArGSerArGLeuAlaArGAlaValAlaGl 529
QY 1085 TCGACCTTACAGCTTCCCTCTCTCC 1112
Db 529 yArGSerAlaGlnProProGlnIlySerPro 538

RESULT 11
ID 022514 PRELIMINARY: PRT: 326 AA.
AC 022514:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE Proline rich protein.
OS Santalum album (white sandalwood).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Santalales; Santalaceae; Santalum.
OX NCBI_TaxID=35574;
RN 11
RP SEQUENCE FROM N.A.
RA Bhattacharya A., Sita L.;
RU Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF020261; AAB70928.1;
SQ SEQUENCE 326 AA; 33407 MW; B10011C45A86DCD7 CRC64;

Alignment Scores:
Pred. No.: 2,43e-05 Length: 326
Score: 150.00 Matches: 63
Percent Similarity: 39.38% Conservative: 26
Best Local Similarity: 27.88% Mismatches: 76
Query Match: 3.33% Indels: 61
DB: 10 Gaps: 11
US-09-806-276a-4 (1-2546) x 022514 (1-326)

```

```

QY 494 CCTGATGATGATGATCCCGCTTGAGAGACTTTGGCCACACCATTCACATTTGCGC-----547
Db 9 ProAspAsnValPheProPro-----AspArgCysThrSerValArGLeuIlySerIlyS 26
QY 548 AAGAGGGCCAGCAGCAGATGCTTTGGAGAAATTTCTTTCTTCAAGGAAATCAGAGCT 607
Db 27 IlyArGlyGlnSerProAla-----33
QY 608 TTACCTGGAACACAGAGAAAGCAGACCTGGCCAGTCTC-----CTGGCATATAGTTC 664
Db 34 --Pro-AspSerArGAsnIlySAsnValIlySerSerGlyIleSerLeuLeuProPr 52
QY 665 TTCCGGGCCAAGACACCTCTGCACTGTG-----TGTTACAGAAAGCGCTCCCGGTC 718
Db 52 oSerAlaProThrThyProThrProLeuGlnValIleSerGlnSerArG-----68
QY 719 CTCAAAATGCTCATCTCCCTCCCGACCATTTGAGAGATCCCAAGCTCTCATGTTGCCCTTA 778
Db 69 -ThrIlySleProSerProThrLeuGlnIlyLeuProPheProLeuAsp-----84
QY 779 TTGTCACCAATGATTTATTTCAACAGAGAGTCTTGGGGCCAGCAAGCTGCGCCAGC 838
Db 85 -----ProArGlySProProSerIlyProPheGlySAsnProAs 97
QY 839 CTTAGCTGGAGCCCGCTCATGAGGAAAGCTCAGGAGAAAGCAGTCTGTGGAGAAAT 898
Db 97 nSerProAlaSer-----LeuIleProThrPArGAr 107
QY 899 GGGACAGCTGCACC-----AGGAGACACTCTGTGGGGC 931
Db 107 gTTpAspHsThrProProLeuPheAsnProProSerAspProProThrProIlyGln 127
QY 932 TCACCGGTTTCATCTCATGTCAGTCCAGGAGGAGAGACAGCAGTCTG-----CC 982
Db 127 rProProThrValArGArGSerProSerProProGlnIlySerAlaSerProSerPr 147
QY 983 AGCTGTCCGAACGATCCCGAGCTGGCCAGCCGAGAGATTTTGACATCCACCCCA 1042
Db 147 oGlyProProAlaAlaArGThrProThrProArGProSerSerProThrSerThrProPr 167
QY 1043 TCGAGCTCATCAAGGAGAAAGTCAAGTCAAGAGAGTCCCTGACCTTACAGTTCC 1102
Db 167 rAspGlyGlnThrProArGProProGlyProAlaIlySerProSerProSerProGlyProPr 187
QY 1103 CTCCTCTCCCGCAGC 1118
Db 187 oSer-----CysSer 190

RESULT 12
ID 09P206 PRELIMINARY: PRT: 1044 AA.
AC 09P206:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE KIAA1522 protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RA MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human
RT genes.XVII.The complete sequences of 100 new cDNA clones from brain
RT DNA Res. 7:143-150(2000).";
RT which code for large proteins in vitro.";
DR EMBL; AB040955; BAA96046.1;
DR InterPro: IPR002965; Pfam: Pf01217; PRINTS: PR01217; PROSITE: PS01217;
DR PRINTS: PRO1217; PROSITE: PS01217;
FT NON_TER 1

```

SQ SEQUENCE 1044 AA; 107934 MW; 5508A7E80E667A44 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.000161 Length: 1044  
 Score: 143.00 Matches: 120  
 Percent Similarity: 34.29% Conservative: 48  
 Best Local Similarity: 24.49% Mismatches: 177  
 Query Match: 3.18% Indels: 147  
 Gaps: 24  
 DB: 4  
 US-09-806-276a-4 (1-2546) x Q9P206 (1-1044)  
 QY 3 GGGCTAGCCCGGAGACCCGCGCCGCTTGCACCGCTCTCGAGCGCAT 62  
 Db 515 GlyGlySerArgArgPro-----ProArgSerProGluArgThrLeu----- 528  
 QY 63 AATGCGGTAGCAGCACCACCGCGGAGCTCGGTCGAGTCGACGCGCAGGGA 122  
 Db 529 SerProSerSerGlyTyrSerSerGlnSerGlyThrProThrLeuProProGly 547  
 QY 123 TTGCGTGGCGCGCGCGGAGCGCGG-----CAGCGGCTGGGA 161  
 Db 548 LeuAlaGlyProProAlaSerProGlyLysAlaGlnProProLysProGluArgValThr 567  
 QY 162 TCCTCAGCGCGCGCGCGCTTTCCTGCTGTGTCACAGCTGATGATGATGCTCT 221  
 Db 568 SerLeuArgSerProGlyAlaSer-----ValSerSerSer 579  
 QY 222 CTAGAGAGCCCTC-----ACTTGGCCGTAACCTCAGGAGGTTCT 260  
 Db 580 LeuThrSerLeuGlySerSerSerSerAspProAlaProSerAspArgSerGlyProGln 599  
 QY 261 -----CTTGGACCCCAT-----CTC 275  
 Db 600 IleLeuThrProLeuGlyAspArgPheValIleProProHisProLysValProAlaPro 619  
 QY 276 ATTTCGAGCCGACTTTCGAGCCACTTGCAGAAAATGAT----- 314  
 Db 620 PheSerProProSerProSerLysProArgSerProAsnProAlaAlaProAlaLeuAla 639  
 QY 315 -----GTGACAGTTCCATCAAA-----AAG 335  
 Db 640 ProAlaValAlaProGlyProAlaSerThrThrAspAlaSerProGlnSerProProThr 659  
 QY 336 GATTGAGAACATATACATCTGTCAAGAAAGTGCCCTTTCGCCGCTTGCAGAAATPAA 395  
 Db 660 ProGlnThrThrLeuThrProLeuGlnGlnSerProValIleSerLys----- 675  
 QY 396 CATTCGCAAAATTCGCAAAATTCGCAAGCCCAATTTACTGAAAGCAGCCAAATPAA 455  
 Db 676 AspGlnSerProProSerProProSerProProSerLysPheSerProProProProThr 694  
 QY 456 AGAAGAGAAATTTAACTGAGGAGCATTTCTCTCGATATGATCATGCTCCCGCG 515  
 Db 695 LysLysProGluValValGlnAlaProSerLysSerGlnThrAlaGlnGluProLe 714  
 QY 516 TTGGAGACTTTCGCCACACATTCGCAAGGAGGAGGAGC----- 560  
 Db 714 u-----GlnAspProAsnTrpProProProProProAlaProGluGln 729  
 QY 561 -----ACGATGCTTTGGAGATATTCTCT-----TCTTC 590  
 Db 729 uGlnAspLeuSerMetLaspPheProProProGluGlnAlaPheSerValAlaSe 749  
 QY 591 -----AAGGAACTAGACCTTTACTGGAACCCAGAGAAACA 631  
 Db 749 rProGluProAlaGlyProSerGlySerProGluLeuValSerSerProAlaAlaSerSe 769  
 QY 632 CACCTGGGCGAGTTCCTGGCGCATATAGATTCTTC-----GGGC 673  
 Db 769 rSerSerAlaThrAlaLeuGlnIleGlnProProGlySerProAspProProAlaLaf 789  
 QY 674 AACAGCAGCTCGACTCTGTGTCA-----CAGAAACG 706

Db 769 oProAlaProAlaProAlaSerSerAlaProGlyHisValAlaLysLeuProGluLysGln 809  
 QY 707 CCCTCCCGGTGCTCAAAATGCCATCTCCCTCCGA-----CCATTGGA 751  
 Db 809 uProValGlyLysSerLysGlyGlyProProArgGluAspValGlyAlaProLeuVa 829  
 QY 752 GATCCCAAGCTCTCATGTTGCCCTTATTGTACACAGTGACATTTAATTCAAACGAG 811  
 Db 829 lThrProSerLeuLeu-----GlnMetValArgLeuArgSe 841  
 QY 812 TCCTTGGGCGCAGCAAGCTGCGGCTTACCTGAGCCGCTCATGAGCAAAAGCT 871  
 Db 841 rValGlyAlaProGlyGlyAlaProThrProAlaLeuGlyProSerAlaProGlnSerP 861  
 QY 872 CAGGAGAAAGCAGCTGTGTTGAGATGGACAGTCACGAGGAGACACCTGTGGGCG 931  
 Db 861 oLeuArgArgAlaLeu-----SerGlyArgAlaSerProValProAl 875  
 QY 932 TCACGCGGTT-----CTGCATCTCAGT-----CCAGCAGAGCAGAGAC 970  
 Db 875 aProSerSerGlyLeuHisAlaValArgLeuLysAlaCysSerLeuAlaAlaSerGln 895  
 QY 971 AGCCACTCTCCAGCCTGTCCGACAGTACCCGACGCTGGCCAGCAGCATGTTTGC 1030  
 Db 895 uGlyLeuSerSerAlaGlnProAsnGlyProProGluAlaGluProArg----- 911  
 QY 1031 CATCCCAACCCCAT-----CGAGCTCATCAAGGAAAGCACTAAGTCAGAGAGTCCCTC 1084  
 Db 912 -----ProProGlnSerProAlaSerThrAlaSerPheIlePheSerLysGlySerArgLys 930  
 QY 1085 TCTGACCTTACAGGTCCTCCCTCTCTCCCTG-----CAGCTTGATCTGGGCGCTCACTTT 1140  
 Db 930 sLeuGlnLeuGlnArgProValSerProGluThrGlnAlaAspLeuGlnArgAsnLeuVa 950  
 QY 1141 GGATGAGGCTCGAATGATGATGATAA 1168  
 Db 950 lAlaGluLeuArgSerLysSerGlnGln 959  
 RESULT 13  
 Q9UF83 PRELIMINARY: PRT; 580 AA.  
 AC Q9UF83;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE Hypothetical 61.1 kDa protein (Fragment).  
 GN DKF2P434C196.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 NC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL133561; CAB63715.1;  
 DR InterPro: IPR002965; P-rich\_extensn.  
 DR PRINTS: PR01217; PRICHEXTNSN.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 580 AA; 61089 MW; C46B82F050F06547 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.000512 Length: 580  
 Score: 137.00 Matches: 104  
 Percent Similarity: 34.32% Conservative: 35  
 Best Local Similarity: 25.68% Mismatches: 159  
 Query Match: 3.04% Indels: 107  
 Gaps: 21  
 DB: 4  
 US-09-806-276a-4 (1-2546) x Q9UF83 (1-580)

```

QY      5  GCTAGCCGAGACCCGCGCCGCTGGGCGGCTTCAGCCGCTTCGAGCGGATTA 64
Db      85  AlaserProthraArgLysProProAlaserProAlaserProthraArgLase----- 102
QY      65  TCGGCTGAGCAGCACCAGCCGCGCAGACTCGGCTGATCGCGACAGCGGAGGATT 124
Db      103  -----SerProthraArgLysProArg----- 110
QY      125  GCGTCGCGCCCGGCGGAGCCGCGGCTGGGATCCTCAGCGCGCGGCTTGTG 184
Db      111  AlaserPromelGlySerProhisArgLaserPromelArgThrProArg----- 128
QY      185  CTGCTGTGTCAACACTGATGATGATGCTCTCTAGGAAGCCTCACTGGCGCTA 244
Db      129  -----AlaserProthrgly-----ThrProSerThraAlaser 139
QY      245  ACCTCAGGAAGGTTCTTTGACCCCATCTCATTTGCAACCCACTTCTGAGCCACTTGA 304
Db      140  Prothrgly-----ThrProSerSerAlaserProthrglyThrProProAr 155
QY      305  GAAATATGATGACAGTCTCTTATCAAAAAGATGCAAAACATATACCATCTGTGAGA 364
Db      155  galaserProthrglyThrProProArgAlaTrrPala-----ThrArgSerProse 172
QY      365  AAGTGGCCCTTCTCCGCTTGCAAAATAGACATTCCTCAATTCGCAAAATGCGCAAG 424
Db      172  rThraLaserLeuThraArgThrProSerArgAlaser-----LeuThraArgThrProProAr 191
QY      425  ACCCATATTACCGAAGCAGCAGCAATTAACAGAAAGAAATTAATACTGAGGAC 484
Db      191  galaserProthraArgThrProPro-----ArgLysSer----- 202
QY      485  ATTCTGTCTCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 544
Db      203  -----ProArgLysSerThraArgLaseProthraArgThr 214
QY      545  GCGCAAGAGGCGCAGCAGCATGCTTTGAGAGATATTCCTTCTTCAAGGAGACTAGAG 604
Db      214  rProProAlaserProthraArgArgPro-----ProArgLaserProthraArgThr 232
QY      605  CTTTACCTGGAACAGCAGCAAGCAACACCTCGGCGGCTTCCTGGGCAATTAAGATTC 664
Db      232  rProProAlaserProthraArgThrSerHisArgLaserProthraArgMetProProAr 252
QY      665  TTCGCGGCGCAAGCAGCAGCTCGACTGTGTTCACAGAAAGCGCCGCGCGGCTCAAA 724
Db      252  galaserProthra-----ArgArgProProAlaserPro 264
QY      725  AATGCCATCTCCCTCCGA-----CCATGAGAGATCCCAAGCTCTCAATGTCCTTA 778
Db      264  oThrglySerProProAlaserPromelThrProProAlaser-----Pro-- 281
QY      779  TTGTACACAGATTAATTATTCCAACAGAGAGCTCTGGGCGCAGCAAGTGGCCAG 838
Db      282  -----ArgThrProProAlaserProthraThrThrProSerArgAlase 297
QY      839  CTTAGCTCGAGCCCGCTCATGAGGAAAGAGCAG-----AGAAAGCAGTCTGTTG 892
Db      297  rLeuThraArgThrProSerThraAlaserProthraThrProSerArgAlaserLeu 317
QY      893  GAGAAATGGAGACGTC-----ACGAGGAGACACCTCGTGGGCTTCAGCGGCTTCT 943
Db      317  tLysMeGlySerThraValSerLeuThraArgThrProProAlaserPro----- 334
QY      944  GCATCTCAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1000
Db      335  -----ThrglyThrProSerArgAlaserProthrglyThr 346
QY      1001  CCGG-----ACTGGCAGCG-----AGCAGCATGTTGACCATCCAGCCCA 1042
Db      346  rProSerArgAlaserLeuThrglySerProSerArgAlaserLeuThrglyThrProse 366

```

```

QY      1043  TCGAGACTCA-----TCAGGGA 1060
Db      366  rArgLaserLeuLeuGlyThrProSerArgAlaserLeuLeuGlyThrProSerArgAl 386
QY      1061  AAGACTAGTACAGAGAGTCCCTCTCTGACCTTACAGGTTCCTCTCCCTCAGCTT 1120
Db      386  aserLeuThrglyThrProProAlaserLeuThrglyThrSerSerThraAlaserLe 406
QY      1121  GATCTTGCGCCCT 1133
Db      406  uThraArgThrPro 410

RESULT 14
09QZR8  PRELIMINARY;  PRT;  2766 AA.
ID  09QZR8
AC  09QZR8
DT  01-MAY-2000 (TREMBLrel, 13, Created)
DE  01-MAY-2000 (TREMBLrel, 13, Last sequence update)
DE  01-JUN-2002 (TREMBLrel, 21, Last annotation update)
DE  PAFIN.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RX  MEDLINE=20449076; PubMed=10896674;
RA  Deguchi M., Iizuka T., Hata Y., Nishimura W., Hirao K., Yao I.,
RA  Kawabe H., Takai Y.;
RT  "A novel multiple PSD-95/Dlg-4/20-1 protein interacting with neural
RT  plakophilin-related armadillo repeat protein/delta-catenin and
RT  p0071."
RL  J. Biol. Chem. 275:29875-29880(2000).
DR  EMBL: AF169411; AAD55940.1; -.
DR  HSP: Q14005; 1116.
DR  InterPro: IPR000194; ATPase_a/bcentre.
DR  InterPro: IPR000719; Euk_Pkinase.
DR  Pfam: PF00595; PDZ; 5.
DR  SMART: SM00228; PDZ; 6.
DR  PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR  PROSITE: PS00106; PDZ; 6.
DR  PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
SQ  SEQUENCE 2766 AA; 293887 MW; 1DC34DF7B46D1DD CRC64;

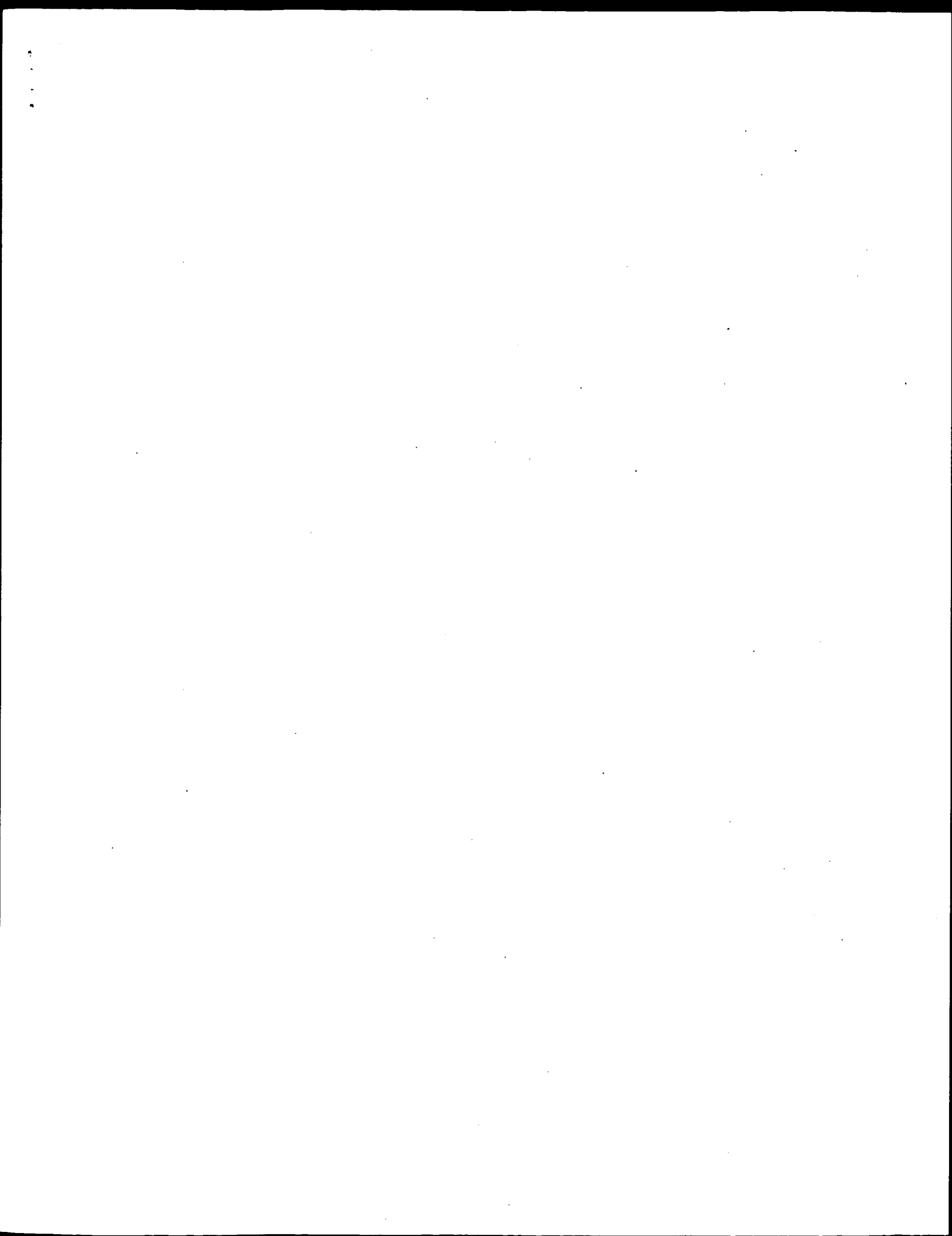
Alignment Scores:
Pred. No.: 0.00126
Score: 135.00 Length: 2766
Percent Similarity: 34.18% Matches: 97
Best Local Similarity: 24.56% Mismatches: 38
Query Match: 3.00% Indels: 128
Gaps: 22

US-09-806-276a-4 (1-2546) x 09QZR8 (1-2766)
QY      103  TCGCGCAGACCGCGCAGGAGTTCGTCGCCGCGGAGGCGCGGCGCAGCGCTGGAT 162
Db      2147  SerSerHisPheGlyArgGlyGlyProSerProHisSerProSerHisSerProHisasp 2166
QY      163  CCTCAGCGCGCGCGGTTTCTCTGCTGTGTGTCAGACTGATGATGATGATGATGATGAT 222
Db      2167  ProGln-----ValProAlaMetGly-GlyLys-----LeuSe 2177
QY      223  TAGGAGCCTTCACTTGGCGCTTAACCTCAGGAGAG-----TTCCTTTGACCCCATCTCA 276
Db      2177  rGluLysThraAlaLysGlyValThrasnGlyngLysValIYrSerValLysPro-LeuL 2197
QY      277  TT--TCGAGCCACTTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 320
Db      2197  eugLthrSerLysasnLeuSerProValaspGlyArgaspAlaSerLaspProLut 2217
QY      321  -----GTTCTATCAAA--AAGATTTCAGAAACATATACCATCTGTGAGAAA 366

```







OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 18, 2003, 17:15:57 : Search time 76.0000

(without alignments)  
6363.010 Million cell updates/sec

```
Title: US-09-806-276A-4  
Perfect score: 4500  
Sequence: 1 cgggctagcccggaaccc...cacttannnnnnnn
```

Scoring table: BLOSUM62

Ygapop	10.0	,	Xgapext	0.5
Ygapop	10.0	,	Ygapext	0.5
Fgapop	6.0	,	Fgapext	7.0
Delop	6.0	,	Delext	7.0

Searched: 283224 seqs, 96134422 residues

Minimum DB count:	566448
Total number of hits satisfying chosen parameters:	566448

```
Maximum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

Command line parameters:  
-MODEL=frame+030-03-1

```

Command line parameters:
-MODEL=frame+ n2p model -DEV=xlp
-o/cg92.1/USPTO.spool/US09806276/runat_18072003_160925.12767/app_query.fasta_1.385
-db=PIR.73 -QEXT=fastan -SUFFIX=ppr -MIMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-units=bits -START=1 -END=1 -MATRIX=blsunn62 -TRANS=human40.cdi -LIST=45
-DOCCALIGN=200 -THR SCORE=post -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-user=US09806276 -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-no_mmap -LARGEOBRY -NEG_SCORES=0 -WAIT -DSPIC=1 -NCPu=6 -ICPU=3
-dev.TIMEOUT=120 -WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -LONGLOG
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

```
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1336	29.5	254	2	T46465	hypothetical prot
2	1305	4.3	391	2	A42973	serum protein MSE
3	137	3.0	580	2	T43481	probable mucin DK
4	136	2.8	1188	2	A49915	extensin-like prot
5	122.5	2.7	3530	2	A59566	unconventional my
6	122	2.7	292	2	S24169	mucin - rat
7	121.5	2.7	543	2	S35047	mucin Juv7 - huma
8	121	2.7	351	2	S50754	hypothetical prote
9	120.5	2.7	519	2	S23756	modifier3 protein
10	120	2.7	551	2	T39092	hypothetical ser P
11	119	2.6	1006	2	T42731	atrophin-1 related
12	115.5	2.6	992	2	T43676	hunchback-related
13	115	2.6	1475	2	S42718	nuclear pore compl
14	115	2.6	2187	2	T30826	nascent polypeptid

15	115	2.6	2500	1	WMHDE2
14	112.5	2.5	1952	2	T48814
13	112	2.5	394	2	C84905
12	111.5	2.5	596	2	S58106
11	111.5	2.5	731	2	JC7701
10	111.5	2.5	954	2	T19765
9	111.5	2.5	1892	2	T18314
8	110	2.4	628	2	UQ0110
7	110	2.4	2351	1	E2HU
6	109.5	2.4	1001	2	S30385
5	109.5	2.4	1459	2	T32371
4	109	2.4	670	2	F84540
3	108	2.4	431	2	T26871
2	108	2.4	384	2	T33504
1	108	2.4	601	2	S56144
30	108	2.4	753	2	UQ0532
31	107	2.4	839	2	T04859
32	107	2.4	955	2	T00247
33	107	2.4	979	2	T50956
34	107	2.4	1188	2	A49960
35	107	2.4	1561	2	T00248
36	106	2.4	753	2	A45301
37	106	2.4	739	2	T44142
38	106	2.4	5282	2	T03454
39	105.5	2.3	382	2	A53715
40	105.5	2.3	368	2	A44265
41	105	2.3	1216	2	T34101
42	104.5	2.3	284	2	T33860
43	104.5	2.3	349	2	T05857
44	104.5	2.3	551	2	UA0851
45	104.5	2.3	3163	1	U01895

## ALIGNMENTS

bad date

## RESULT 1

C:Species: Homo sapiens (man)  
C:Date: 04-Feb-2000  
C:Accession: 146465  
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000  
A:Reference number: 223034  
A:Accession: 146465  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-254 <AAA>  
A:Cross-references: EMBL:AL136842  
A:Experimental source: adult testis; clone DKFzp34A0530  
C:Genetics:  
A:Note: DKFzp34A0530.1

Alignment Scores:  
Prod No.

Seed, NO.:	5,29e+102	Length:	25
Score:	1326.00	Matches:	254
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	29.47%	Indels:	0
DB:	2	Gaps:	0

1-2546 (1-2546) X T46465 (1-2546)

.....CCCAATTTACCTGAAAGCAGCCAATAACAAGAAGGAAGAAATT 472

...LysAlaAsnLysGlyLysPhe 20

473 ...

OV

[illegible][illegible]

100

[illegible][illegible]



submitted to the Protein Sequence Database, September 1999  
 A:Reference number: 218723  
 A:Accession: T11264  
 A:Molecule type: mRNA  
 A:Residues: 262-580 <POU2>

A:Cross-references: EMBL:AL117481; NID:95911958; PDB:1CAB5954.1; PDB:95911959  
 A:Experimental source: adult testis; clone DKFZp434B061  
 C:Genetics:  
 A:Note: DKFZp434C196.1; DKFZp434B0635.1; DKFZp434B061.1

## Alignment Scores:

Pred. No.: 0.00493  
 Score: 137.00  
 Percent Similarity: 34.328  
 Best Local Similarity: 25.688  
 Query Match: 3.048  
 DB: 2  
 Gaps: 21

US-09-806-276a-4 (1-2546) x T43481 (1-580)

```

QY 5 GCTAGCCCGGAGACCGCGCCGCTTGCAGCGCGCTTGCAGCGCGTCTGAGCGGATTA 64
DB |||||
DB 85 AlaserProthArgLysProProArgAlaserProArgThProSerArgAla----- 102
QY 65 TCGCGTGAGCAGCAGCAGCGCGCGAGCTGCGATCTGCGCAGCAGCGGAGGATT 124
DB |||||
DB 103 -----SerProthArgLysLeuProArg----- 110
QY 125 GCGTGGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTC 184
DB |||||
DB 111 AlaserProMetCysSerProHisArgAlaserProMetArgThProProArg----- 128
QY 185 CTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 244
DB |||||
DB 129 -----AlaserProthArgLys-----ThProSerThAlaser 139
QY 245 ACCTCAGCAGAGGTTCTCTTGAACCCATCTCATTTGCAAGCAGCTTGAAGCGCATTGA 304
DB |||||
DB 140 ProthArgLys-----ThProSerThAlaserProthArgLysProProArg 155
QY 305 GAAATATGATGTGACACTTCCATATCAAAAGATTCGAAACATATACCATCTGTGAGA 364
DB |||||
DB 155 GalaserProthArgLysProProArgAlaTIPala-----ThArgSerProse 172
QY 365 AAGGCGCGCTTCTCCCGCTTGCAGAAATAGACATTTCTCAAAATGCGCAGCAAG 424
DB |||||
DB 172 TThAlaserLeuThArgThProSerArgAlaser-----LeuThArgThProProArg 191
QY 425 ACCCAATTATACCTGAAGCAGCAGCAATTAAGAAAGAAAGAAATTAACCTGAGGAC 484
DB |||||
DB 191 GalaserProthArgThProPro-----ArgLysUser----- 202
QY 485 ATTCTGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 244
DB |||||
DB 203 -----ProArgMetSerHisArgAlaserProthArgTh 214
QY 545 GGCAGAGAGGCGCAGCAGATGTTTGGAGATATTTCTTCTTCAAGGAGCAGTACAG 604
DB |||||
DB 214 ProProArgAlaserProthArgArgPro-----ProArgAlaserProthArgTh 232
QY 605 CTTTACTGGAACACGAGGAGAAAGCAGCAGCTGGCGGCGGCGGCGGCGGCGGCGG 664
DB |||||
DB 232 ProProArgLysLeuThArgThProSerArgAlaserProthArgMetProProArg 252
QY 665 TTCGGGCGCAGCAGCAGCTGCTGTTCAGAGAAAGCGCGCGCGCGCGCGCGCGCG 724
DB |||||
DB 252 GalaserProth-----ArgArgProProArgAlaser 264
QY 725 AATGCCATCTCCCTCCCGA-----CCATTGGAGATCCCAAGCAGCTCATGTGCGCTTA 778
DB |||||
DB 264 CThArgLysProProArgAlaserProMetThProProArgAlaser-----Pro-- 281
QY 779 TTGTACAGAGTACATTATTCAGAAAGAGAGTCTTGGCGCGCAGCAAGCAGTGGCGAG 838

```

```

DB 282 -----ArgThProProArgAlaserProthArgThProSerArgAlase 297
QY 839 CTTAGCTCGAGCGCGCGCTCATGAGGAGAAAGCTCAGG-----AGAAAGCAGCTGTG 892
DB |||||
DB 297 LeuThArgThProSerThProArgAlaserProthArgThProSerArgAlaser 317
QY 893 GAGAAATGGAGACAGTCC-----ACCAGGAGACAGCTGTGCGGCGCTCCAGCGGCTT 943
DB |||||
DB 317 TysMetCysLeuThArgThAlaserThArgThProProArgAlaserPro----- 334
QY 944 GCATCTCAGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1000
DB |||||
DB 335 -----ThArgLysProSerArgAlaserProthArgTh 346
QY 1001 CCGG-----ACTGGCAGCGCG-----AGGACATGTTGACCATCCACCCCA 1042
DB |||||
DB 346 ProSerArgAlaserLeuThArgLysProSerArgAlaserLeuThArgThProse 366
QY 1043 TGGCAGCTCA-----TCAGAGGA 1060
DB |||||
DB 366 ArgAlaserLeuThArgLysProSerArgAlaserLeuThArgThProse 386
QY 1061 AAGACTAAGCAAGAGAGAGTCCCTCTGACCTTACAGGTTCCCTCTCCCTCAGCTT 1120
DB |||||
DB 386 aserLeuThArgThProProArgAlaserLeuThArgThProSerThAlaser 406
QY 1121 GATCTTGGCGCGCT 1133
DB |||||
DB 406 ThArgThProPro 410

```

RESULT 4  
 S49915  
 C:Species: Zea mays (maize)  
 C:Date: 05-Mar-1995 #sequence\_revision 12-May-1995 #text\_change 29-Oct-1999  
 R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.  
 submitted to the EMBL Data Library, June 1994  
 A:Description: Pex genes: Pollen-specific genes with extensin-like domains.  
 A:Accession: S49915  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1188 <RUB>  
 A:Cross-references: EMBL:Z5465; NID:9600117; PDB:1CAB4230.1; PDB:9600118

## Alignment Scores:

Pred. No.: 0.0453  
 Score: 126.00  
 Percent Similarity: 34.198  
 Best Local Similarity: 23.008  
 Query Match: 2.808  
 DB: 2  
 Gaps: 12

US-09-806-276a-4 (1-2546) x S49915 (1-1188)

```

QY 240 CCGTAACTCAGAGAGTCTCTTGAACCCATCTATTGAGAGCAGTCTGAGAGCA 299
DB |||||
DB 696 ProProArgLysLeuThArgThProProSerThProSerLysProProSerPro 715
QY 300 CTTGAGAAATATGATGTGACAGTTCCTATCAAAAGATTCAGAAACATATCCATCTGT 359
DB |||||
DB 716 GluLysProSerProProLysGluProValSerSerProProGlnThr----- 731
QY 360 GAGAGAAAGTGCCCTTCTCCCGCTTGCAGAAATATGACATTCGAAATTCAGAAATGCGCAG 419
DB |||||
DB 732 -----ProLysSerSer 735
QY 420 CCAAGACCCCAATTACTGGAAGCAGCAGCAATTAACAGAAAGAAAGAAATTAACCTGA 479
DB |||||
DB 736 ProProPro-----AlaProVal----- 741

```



## Alignment Scores:

Pred. No.: 0.0783 Length: 292  
 Score: 122.00 Matches: 56  
 Percent Similarity: 37.45% Conservative: 32  
 Best Local Similarity: 23.83% Mismatches: 95  
 Query Match: 2.71% Indels: 52  
 DB: 2 Gaps: 10

US-09-806-276a-4 (1-2546) x S24169 (1-292)

QY 519 GAGACTTTCGGACATCATTCATTTGGCAAGAGGCCAGACGATGCTTTGGAGATA 578  
 DB 48 GlutProSerProProSerThrIleSerGluThrAlaSerSerThr----- 64  
 QY 579 TTTCTTCTTCAAGGAGATACAGCTTTTACCTGGAAACGAGGAAAGACACACCTGG 638  
 DB 65 ---ProThrThrGluSerThrSerThrSerThrSerThrValProThrProThr 83  
 QY 639 GCCAGTTCCTGGGCAATATGAGTCTCCGGGCAACAGACCTCGACTGTGTCA 698  
 DB 84 ProThrThrIleGluSerGluThrProThrSerThrValProThrThrGlySerThr 103  
 QY 699 CAGAAAGCCCTCCCGGCTGCTCAAAATGCCA-----TCTCCCTCCGA 743  
 DB 104 SerSerThrProThrThrGlySerThrProThrSerProSerThrProSerPro 123  
 QY 744 CCATTGGAGATCCCAAGCTCTCATGTCCTTATTGTCACAGTACATTTAATTCGA 803  
 DB 124 ProThrGluSerThrThrLeuSerThrPro-----ValThr 136  
 QY 804 AACGAGTCTCTCGGCGCCAGCAAGCTCCAGGCTTAGTCGAGCCCTCATGAGG 863  
 DB 137 ThrThrAlaThrSerThrThrSerSerProGly---ThrThrSerProHeAlaThr 155  
 QY 864 AAAAGCTCAGGAAAGAGCTGTGTGAGATGAGACAGTACACAGGAGAGACACT 923  
 DB 156 SerSerVal-----SerSerThrProProSerPro 165  
 QY 924 ---CGTGGGCTCCAGCGGTTCTGCATCTCAGTCAGCCAGCCAGAGACAGCC 974  
 DB 166 ProSerSerThrSerThrGlyProThrThrSerSerGlyMetProThrSerThrThr 185  
 QY 975 ACTCTCCAGCCGTCGGAAGATACCCG-----ACTGGCAGCCAGACATGTTTG 1028  
 DB 186 ThrGlyProThrSerProThrThrThrProProSerThrSerThrPro 204  
 QY 1029 ACCATCCACCCCATCGAGCTCA-----TCAGGAGAAAGACTTAAGTACAGAGGCC 1082  
 DB 205 ThrValProThrSerThrThrGluAlaIleThrGlnThrAlaGlyLeuSerThrThrPro 224  
 QY 1083 TCTCTACTTACAGTTCCCTCCCTCCCTGCGACAGTTGATCTGGGCCCTCACTTTGG 1142  
 DB 225 ThrMetGluThrThrAlaThrSerSer----- 233  
 QY 1143 ATGAGTGCTGAATGATTAATAAATAGTACAGATGCCA 1187  
 DB 234 -----TrrGlyThrValSerSerAlaSerPro 242

## RESULT 7

S35047  
 muscin JUL7 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 10-Dec-1993 #sequence\_revision 10-Nov-1995 #text\_change 08-Oct-1999  
 C:Accession: S35047  
 R:Dufose, J.; Porchet, N.; Audie, J.P.; Guyonnet Duperat, V.; Laine, A.; van-Seuningen,  
 Biochem. J. 293, 329-337, 1993  
 A>Title: Degenerate 87-base-pair tandem repeats create hydrophilic/hydrophobic alternati  
 A:Reference number: S35047; MUID:93343858; PMID:7916618  
 A:Accession: S35047  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-543 <DUF>

A:Cross-references: EMBL:X74370; NID:9407081; PIDN:CAAS2408.1; PID:9407082  
 A>Note: the authors translated the codon AAA for residue 63 as Gln and CGG for residu  
 A>Note: the amino acid sequence from Fig. 2 is inconsistent with the nucleotide seque

## Alignment Scores:

Pred. No.: 0.0949 Length: 543  
 Score: 121.50 Matches: 89  
 Percent Similarity: 31.13% Conservative: 43  
 Best Local Similarity: 20.99% Mismatches: 157  
 Query Match: 2.70% Indels: 135  
 DB: 2 Gaps: 21

US-09-806-276a-4 (1-2546) x S35047 (1-543)

QY 99 TGAATTCGCGACAGCGGAGGATTCGTCGCCCGGAGGCCCGGCGAGCGCTG 158  
 DB 172 TrpIleuThrGluLeuThrThrAlaIleThrThrAlaGlyThrGlyProThrAla 191  
 QY 159 GGATCCTCAGCGGCGCGGCTTGTCTGCTGTGTGTGTCAGACTGAGATGTA----- 212  
 DB 192 ThrProSerThrProGlyThrThrThrIleLeuThrGluLeuThrThrAlaThr 211  
 QY 213 -----ACTGCTCTAGAGAGCTC-----ACTGGCGG 242  
 DB 212 ThrThrAlaSerThrGlySerThrThrAlaThrLeuSerThrProGlyThrThrTrp--- 230  
 QY 243 TAACTCAGAGAGTCTCTTTCAGCCCATTCATTCGAAAGCCACTTGAAGCACTT 302  
 DB 231 -----IleLeuThrGluProSer 236  
 QY 303 GAGAAATGATGTGACAGTCTCTATCAAAAGATTCGAAACATATACATCTGTGA 362  
 DB 237 ThrThrAlaThrValThrValProThrGlySerThrThrAlaIleSerSerThrGlnAla 256  
 QY 363 GAAATGCGCCCTTCTCCCGCTTGCAAAATAGACATTCATTCAAAATGCCAGCA 422  
 DB 257 ThrAlaGlyThr-----ProHisValSerThrThrAlaThr 268  
 QY 423 AGACCCCATTTACCTGGAAGCAGCCATTAACAAGAAAGAAATTAAGTGAAGG 482  
 DB 269 ThrProThrValThrSerSerThrLeuLeuSer----- 281  
 QY 483 ACATTCCTCTCTGATATGATCACTCCCGCTTGAGACTTGGCACACCATCCACA 542  
 DB 282 -----ProGlyLeuGln-----LeuProSerAla 289  
 QY 543 TTGCAGAAAGAGGCC---AGCAGATGCTTTGGAGATTTCTCT---TTCTTC 590  
 DB 290 LeuArgSerThrAlaThrAsnThrHisSerThrGlnLeuThrHisProLeuLeu 309  
 QY 591 AAGGACACTAGAGAGCTTTTACCTGGAACAGAGAAAGACACCTGGCCAGTTCCCTG 650  
 DB 310 ProGlyThrThr-----TrrThrArgLeuSerGlnThrThrProThrAla 325  
 QY 651 GGCATTAATGAGTCTCCGGGCAACAGCAGCCTCGAGCTGTGTGA----- 688  
 DB 326 ThrMetSerThrAlaThrProSerSerThrProGluThrValHisThrSerThrValLeu 345  
 QY 698 ----- 698  
 DB 346 ThrAlaThrAlaThrThrThrGlyAlaProAlaLeuThrProProProProGln 365  
 QY 699 ---CAGAAAGCCCGCCCGGTCGCAAAATGCAATCTCC-----TCCGAGCA 746  
 DB 366 GluGlnLeuThrLeuProGlyCys---ArgLeuProGlnProArgLeuHisSerHisPro 384  
 QY 747 TTGGAGATCCCAAG-----CTCTCATGTTGCCCTTATGTCTACSCAGTGA 791  
 DB 385 LeuLeuGlnProArgAspGlyThrHisAlaSerSerValAspGlnHisAsnHisThr 404  
 QY 792 CAT-----TTAATTCGAAACAGG 809  
 DB 405 HisAsnHisThrHisAsnGlnThrProLeuHisGlyAspProLeuHisProGlyAsp 424



Oy		CTTCTGTAAG--CCACTGTAGAATAATGATGTGCACAGTTCTCTCATTCAAATAAAGATTTCAGAA	344
Oy		:	:
Dd	156	GltLeuLysAspPro-IleArgLys-----	-LysAlaGlyArgGly 168
Oy	345	ACATATPCCCACTCTGTGAAGAAATGSCCCTTCTCCCGCTTGCAAAAATAGCATTTTCAA	404
Dd	168	sProlenPrProGluInsLysAlaIalaArgHtProValSerLeuAlaLysValLeuIIy	188
Oy	405	ATTCCAAAATGCCAGCCAAAGCCCCCAATTTTACCCTGAAGACCCAAATTAACAAGAAAGAA	464
Dd	188	S-----	-ThrThrArgLy 192
Oy	465	AGAAATTTAAACSTGAGSGAGCATTTCTGTCTCCATGATATCACAGTCCCCTGGTAGAGACT	524
Dd	192	sAspLeuIIyThrserAlaIalalysLeu-ProrProIleuSerAlaProValAlaIagLy-	211
Oy	525	TTCGCCACAA-----	-CAATCCACATTTGGCAAG 551
Dd	212	LeuAlaIaleuLysAlaHisThrLysGluaIacysLysLysIleProserThrMetAlaThr	231
Oy	552	AGGGCCACAGCATGTCTTGTGAATATTTCTTTCTTCAAGGAGACTACGACTTTTAC	611
Dd	232	ProGluasnLeuAlaSerLeu-----	-MetLysGlyMetAlaGly----- 244
Oy	612	CTGSAAACACAGGAGAAAGCACACCTGGGCCAGTTCCCTG-----	-GGCATTA 656
Dd	245	---SerProserAlaArgLysIYlIeTrpIlnSerSerIleValaHisTryMeIasnaRgmeL	263
Oy	657	ATGACTTCTTCGGGGCCAACAGCCACCTGGGCTGTGTTCACAGAAACCCCTCCCGG	716
Dd	264	SerGIlnSerGlnValGlnAlaIalaSerArgLeuAlaLeuLysAlaGlnAlaThrAsnLys	283
Oy	717	TGCTCA---AAATGCCATCTCCCTCCGACCATTTGAGAGATCCCAAGCTCTCATGTGC	773
Dd	284	CysGlyLysLeuLysAspLeuLysValaArgThrGlnLysGlyLysLeuGlyLysSer	303
Oy	774	CTTATTTGCACACAGCATTTAATTCCAACACAGAGACTCTTCGGGCGACAA-----	827
Dd	304	ProAlaGlyLysValaProLysAlaProGlyLysGlyAlaIaGlnGlnAlaGly	323
Oy	828	-----AGTGGCCAGGCTTAACTGCGGAGCCGCTCATGGAGAGAAAAAGCTCAGAGGA	878
Dd	324	AsnHisSerGlySerProGlyAlaGlnLeuAlaProThrGlnGlnLeuSerLeuGlnVal	343
Oy	879	AAACAGACTCTGTGGAGAATGGAGACAGTCCACGAGGAGACACACTCTGGGGCTTCACGG	938
Dd	344	LeuAspLeu-----	-GlnSerValLysAsnGlyValaProGlyValaGlyLeu 358
Oy	939	GTTCTGCATCTCACTCCAGCCAAAGCCAGACAGCCACTCTCCAGCTCT-----	989
Dd	359	LeuAlaIaRghHisAlaProAlaLysAlaIlaProAlaThrAsnProAlaThrGlyLysGly	378
Oy	990	CCGAACAGTACCOCGAGCTGGCGACGCCAGAGCAATGTTTGACATCTCCACCCCATCGAGAC	1045
Dd	379	ProGlySerGlyProThrGly-----	-AlaAsnMetThrAsnIaIaProThrAspAsn 395
Oy	1050	TCATCAAGGAAAGACTAAGTCAGAGAGTCCCTCTGTG	1088
Dd	396	AsnLysGlyLysLeuThrCysLysAlaThrAlaLeu	408

Alignment Scores:	
Pred. NO.:	0.169
Score:	119.00
Percent Similarity:	33.86%
Best Local Similarity:	23.28%
Query Match:	2.64%
DB:	2
Length:	1006
Matches:	88
Conservative:	40
Mismatches:	143
Indels:	107
Gaps:	14

```

Db      115 rlysthringlu1lleSerArgproAasnerProserGcluglYcUglYluserSeras 135
QY      532 CACCAATCCATTGGCCAAAGAGGGCCACCATGATGTCTTTGGAGATATTTCCCTTCTTTGA 591
       :::          |||||         |||          |||||         ||
Db      135 pSerArgSerValnaAspGlUGlYserSerAsPro---LysaspIleSp-----GI 152
QY      592 AGGCAGACTCGAGCTTTTAACCTGGAAACACAGGAGAAGAACACACCCTGGCCAGTCCTTG 651
       |   |||           |||
Db      152 nAspsnAlaSerThrSerProSerIleProSer-----ProGI 165
QY      652 GCATATAAGTCTTCCGGGCCAACACACCTCGGACTCtGTTTCACAGAAACCCCTC 711
       |||||         |||          :::::        |||          ::|||
Db      165 nAspsnnglu-SerAspSerAspSerSerIaladInglInglmetLeuglnThrglnProp 185
QY      712 CCCGGTGTCAAAMATGCCATCTCCCTCCCGCACCATGGAGATCCCAAGCTCTCATGTT 771
Db      185 ro-----AlaleuglnAlaProSerGclYAlalaaser-----195
QY      772 GCCCTTATTTGTACCACTGACATTTAATTCCAAAAGGAGCTCTCGGCCACGAACCT 831
Db      196 -----AlaProSerThrAlaProProDylThrThrglnL 207
QY      832 GCCCAGCGTTAGCTGGCAGCCCTCATGGAGAAAAAGCTCAGSGAAMAACACTGTGT 891
       |||          ::|||         |||
Db      207 euProThnProGlyProThnProSer-----215
QY      892 GGAGAAATGGAGACAGTCCACAGGAGAACACCTCGTGGGGGTCACGGCTTCTGCAT--- 947
Db      216 -----AlatHrThvalProProGlnglYSerProIalrhSerGlnProp 231
QY      948 -----CTCAGTCCAGCC-----AAGCAGAGA 969
Db      231 roAnsglnHrghInsErThrValAlaPrroIalAlahIsThrleuIlleglnGlnThrPro 251
QY      970 CAGCCACTCCTCCAGCTGTCCGAAACAGTACCSCCGACATGGCCAGCCAGGACATGTTTGA 1029
       |||          |||||         |||          |||||         |||
Db      251 hrleuHisProProAlaHrleuProSerProHisProProlleuGlnPro-----MetTr 268
QY      1030 CCATCCCACCCCATCGCAGCTCATCAGGAAAGACTAAGTCAGAGAGACCTCTCTCTGA 1089
       |||          |||||         ::|||         |||||         |||
Db      268 hralaProProSerGlnAsnSerAlagln-----ProHisProGlnProSerleuH 285
QY      1090 CCTTACAGGTTCCCTCTCTCCCTGCAGACTGATCTTGGCCCT 1133
Db      285 lsglYglnGlyProProGlyProHisSerleuGlnHrghlPro 299

RESULT 12
T43676
hunchback-related protein - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:date: 21-Jan-2000 #sequence,revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T43676
R:Ray, D.S.; Stanley, H.M.; Han, M.; Wood, W.B.
Dev. Biol. 205, 240-253, 1999
A:title: A Caenorhabditis elegans homologue of hunchback is required for late stages
A:Reference number: 222622; MUID:99117349; PMID:9917360
A:Accession: T43676
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-982 <RAY>
C:Genetics:
C:cross-references: EMBL:AF097737; NID:g4323034; PIDN:AAD16170.1; PID:g4323035
A:Gene: hbl-1

Alignment Scores:
Pred. No.:
Score: 0..329 Length: 982
Percent Similarity: 115.50 Matches: 83
Best Local Similarity: 38..25% Conservative: 57
Query Match: 22..68% Mismatches: 146
                Indels: 80
                Gaps: 16

```

```

QY 227 AAGCTCACTTGGCCGTAACCTGACGAAAGTTCTTGTGACCCCATCTATTCGAGCC
  |||||
Db 174 LysProGluSerThrSerValGluGly-----ThrSerSerAspTyrGlnVal 189
QY 287 ACTTGTGAAGCCACTTGAAGAAATGATGTCAGATTCTATCAAAAAGATTCCAGAAC
  |||||
Db 190 ThrSerGluProVal-----
QY 347 ATATACCATCTGTGAAGAAAGTGGCCCTTCTCCCGCTT---GCAAAATGACATTCCTCA
  |||||
Db 195 -----GlnMetProGlnMetProIleProValIleProSerPheLeu 208
QY 404 AATTCAAATGCCAGCAGCAGCCCAATTTAC---CTGAAGCAGCCCAATTAACAGAA
  |||||
Db 209 LysAsnSerLeuProAlaProIleProIleThrProThrGlnSerAlaAsnValGluArg
  |||||
QY 461 GGAAGAAATTTAACTGAGGAGCATCTGTCTCTGATATGATGATGATCCCGCTTG--
  |||||
Db 229 SerAsnSerProSerIleGlnGlnAlaLeu-LeuLeuThrLeuSerGlnGlnGlnPheAl
  |||||
QY 519 -GAGACTTTCGCCACACCATCCACATTTGGCAAGAGGCGCAGCAGATGCTTTGGAGAT
  |||||
Db 248 aglValAlaPheAlaGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla
  |||||
QY 578 ATTTCTTCTTCAAGGAGACTAGAGCTTTTACTGGAACACGAGGAAAGCAGACCTG
  |||||
Db 264 rIleGlyPheGlnArgSerGlyThrSerAlaPheLeuAsnIleGlnProLysGluMetSe
  |||||
QY 638 GGCAGTTCCTCGGCGCATATGACTTCTCCGCGCAACGACACTGCGACTGCTGCTTC
  |||||
Db 284 rMetSerSerAlaAsnAsnAsnAsnGlnGlnAlaProAlaSerThrValSerAlaCysSe
  |||||
QY 688 ACAGAA-----CGCCCTCCCGGCTGCTCAAAAATGCC 730
Db 304 rThrProThrThrThrThrSerAlaSerPheCysAlaPro-ProGlyLeuGlyProValA
  |||||
QY 731 ATCTCCCTCCGACCATTTGAGAG---TCCCAAGCTCTCATGTTGCCCTTATG-----
  |||||
Db 324 LaleuPro-ProThrGlnAsnGlnGlnThrPrometLeuValCysProIleCysGlyPhe
  |||||
QY 782 -----TCACCATGACATTTATTTCCAAACAGAGACTCTTGCGGCGCAGCAAGCTG
  |||||
Db 344 MetCysProSerThrPheHisPheAsnSerHisMetAsnThrHisGlyAspHisGlnCys
  |||||
QY 833 CCCAGGCTTAGCTGGAGCCCGCTCATGAGCAAAAAGCTCAGAGAAAGACAGTCTGTG
  |||||
Db 364 Ser-----MetCysAspTyrThrSerArgThrGlnGlyArgLeuLysHisMetArg
  |||||
QY 893 GAGATGGACAGTCCAC-----CAGGAGACACCTCGTGGGC 931
Db 382 GluSerHisThrValGlnGlnGlnLeuArgAlaGlyPheGlnSerGluProAlaLysGlu
  |||||
QY 932 TCCAGCGGTTCTGCATCTCAGTCCAGCCAGAGACAGCAGCAGCTCTCCAGCTGTGC
  |||||
Db 402 SerAlaSerSerProLysAsnLeuSerLeuSerLysSerLysSerLysSerLysSerLys
  |||||
QY 992 GAACAGTACCCCGCAGCTGCCAGCAGGACATGTTGACATCCACCCCATCGAGCTC
  |||||
Db 420 -----ProIleAsnGlnIlePheAsnLeuSerThrThrMetAlaSer 433
QY 1052 ATCAAGGAGAAAGACTAG---TCAGAGGAGTCCCTCTGACCTTACAGATTC 1102
Db 434 rIleLeuAspSerThrAsnAsnAlaValSerSerThrThrThrGlnGlnProSerAla
  |||||
QY 1103 CTCTCTCCCTCGACGTTGATCTTGGG-----CCCTCACTTTGGATGAGGTG-----
  |||||
Db 454 LeuSerAlaLeuThrLeuAsnMetSerSerThrProSerLeuLeuSerThrLeuAlaHis
  |||||
QY 1151 -----CTGAATGATGATGATAAATAG-----TAACAAGATGCCAAGCTT 1192
Db 474 SerSerPheGlyValSerAlaLeuAspGlnIleLysAlaIleSerGluAsnProSerPhe
  |||||

```

```

QY 1193 TTTCTTTGGGG 1204
Db 494 MetProGluGly 497

RESULT 13
S42718
nuclear pore complex protein nup153 - human
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence,revision 13-Jan-1995 #text_change 05-Nov-1999
C:Accession: S42718; S37477
R:McMorrow, I.; Bastos, R.; Horton, H.; Burke, B.
Biochim. Biophys. Acta 1217, 219-223, 1994
A:Title: Sequence analysis of a cDNA encoding a human nuclear pore complex protein, h
A:Reference number: S42718; MUID:94154002; PMID:8110839
A:Accession: S42718
A:Molecule type: mRNA
A:Residues: 1-1475 <MCN>
A:Cross-references: EMBL:225335; NID:9406224; PID:CA080982.1; PID:9406225

Alignment Scores:
Pred. No.:
Score: 0.385 Length: 1475
Percent Similarity: 115.00 Matches: 88
Best Local Similarity: 36.58* Conservative: 51
Query Match: 23.16* Mismatches: 141
DB: 2.568 Indels: 100
Gaps: 19

US-09-806-276a-4 (1-2546) x S42718 (1-1475)
QY 122 ATTCGCTGCGCCCGGAGGCGCGGCGGCGAGC-----
  |||||
Db 814 ValSerCysMetSerGluLysProGlySerSerValProAlaSerSerSerThrVal 833
QY 155 ---CGTGGATTCCTACAGCGCGCGCGGCTTGTCTGCTGTTGTGCTCAAGCTGGATGATGT
  |||||
Db 834 ProValSerLeuProSerGlyGlySerLeuGlyLeuGlnLysPheLysLysProGluGly 853
QY 212 AACTGG-----CTCTCTAGAGAGCCCTCAC-----TTGGCC 241
Db 854 SerTyrAspCysGluLeuLeuValGlnAsnLysAlaAspSerThrLysCysLeuAla 873
QY 242 GTAACTTCAGGAAGGTTCTTCTTGAACCCCATCTATTCGAAGCC-----ACTTCGAA 295
Db 874 CysGluSerAlaLysProGly---ThrLysSerGlyPheLysGlyPheAspThrSerSer 892
QY 296 GGCACCTGAGAAATGATGATGACAGTTCCTATCAAAAAGCATTCAGAAACATATACAT
  |||||
Db 893 SerSerSerAsnSerAlaAlaSerSerSerThrLysPheGly-----906
QY 356 CTGTGAAGAAAGTGCGCTTCTCCCGCTTGCAGAAATGACATTCATTCGAAATG 415
Db 907 -----ValSerSerSerSerGly 913
QY 416 CCAGCAAGACCCCAATTTACTGTAAGACAGCCCATTAACAAGAAAGAAAGAA-----469
Db 914 ProSerGlnThr-----LeuThrSerThrGlyAsnPheLysPheGlyAspGlnGlyGly 931
QY 470 TTTAACTGAGGAGCATTTCTGCTCTGATATGATGATGATGATGATGATGATGATGATGAT
  |||||
Db 932 PheLysIleGlyValSerSerAspSerGlySerIleAsnPrometSerGluGlyLysPheLys
  |||||
QY 512 -----CCGTTGAGAGCTTTCGCCACACACAC-----CAC 541
Db 952 PheSerLysProIleLysAspPheLysPheGlyValSerSerGluSerLysProGluGlu 971
QY 542 ATTGGCAAGAGGCGCAGCAGCAT-----GCTTTGAGAGATATTTCTTCTTCAAGGG 595
Db 972 ValLysLysAspSerLysAsnAspAsnPheLysPheGlyLeuSerSerGlyLeuSerAsn 991
QY 596 AACTAGACCTTTTCTGCTGGAACCCAGAGAGAACACACACCTCGGCGGCGGCGGCGGCGG
  |||||
Db 992 ProValSerLeuThrProPheGlnPheGlyValSerAsnLeuGlnGlnGlnLysLys 1011

```

QY 656 AATGATGTTCTCCGGGCAACAGCAGCTCGACTCT-----GTGTCCAGAA 703  
 DB 1012 GluIuLeuProIysSerSerSerAlaGlyPheSerPheGlyThrGlyValIleAsnSer 1031  
 QY 704 AGCCCTCCCGGTG-----CTCAAAATGCCATCTCCCTCCG 742  
 DB 1032 ThrProAlaProAlaAsnThrIleValIleThrSerGluAsnLysSerPheAsnLeuGly 1051  
 QY 743 ACCATTGAGGATCCCAAGCTCTCATGTGCTCCCTATTATTCACCAAGTCAATTATTC 802  
 DB 1052 ThrIleGluThrLysSerAlaSerValAlaProPheThrCysLysThrSerGluAlaLys 1071  
 QY 803 AAACAGAG-----TCTTCCGG-----CCAGCAAG 829  
 DB 1072 LysGluGluMetProAlaThrLysGlyLysPheSerPheGlyAsnValGluProAlaSer 1091  
 QY 830 CTGCGGAGCTTACGTGCGAGCCGTC-----ATGAGGAAAGCTCAGAG----- 877  
 DB 1092 LeuProSerAlaSerValPheValIleGluArgThrGluGluLysGluGluGluProAla 1111  
 QY 878 AAAAGCAGCTCTTGAGAAATGGAGTCACAGCCAGGAGACACC----- 922  
 DB 1112 ThrSerThrSerLeuValPheGlyLysValAlaAsnGluGluProLysCysGluPro 1131  
 QY 923 -----TCTGGGGCTCCAGCGGTTCTGCATCTCAGTCACGACGAGGAGAGACGCCAC 976  
 DB 1132 ValPheSerPheGlyAsnSerGluGluThrLysAspGluAsnSerLysSerThrPhe 1151  
 QY 977 TCTCCAGCCG-----TCCGAAAGTACCCGACTGCCAGCCGAGAGACATGTT 1027  
 DB 1152 SerPheSerMetThrLysProSerGluLysGluSerGluGluProAlaLysAlaThrPhe 1171

## RESULT 14

T30826  
 nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse  
 N:Alternate names: alpha-NAC protein  
 C:Species: Mus musculus (house mouse)  
 C>Date: 22-Oct-1989 #sequence\_revision 22-Oct-1999 #text\_change 24-Nov-1999  
 C:Accession: T30826  
 R:Yotov, W.V.; St-Arnaud, R.  
 Genes Dev. 10: 1763-1772, 1996  
 A:Title: Differential splicing in of a proline-rich exon converts alphaNAC into a muscle  
 A:Reference number: Z20889; MUID:96312450; PMID:8698236  
 A:Accession: T30826  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 12187 <YOT>  
 A:Cross-references: EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAB18732.1  
 C:Genetics:  
 A:Gene: Naca  
 A:Map position: 10  
 A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3  
 A:Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ  
 C:Keywords: alternative splicing; DNA binding; transcription factor

Alignment Scores: 2187  
 Pred. No.: 0.41 Length: 87  
 Score: 115.00 Matches: 87  
 Percent Similarity: 32.99% Conservative: 42  
 Best Local Similarity: 22.25% Mismatches: 148  
 Query Match: 2.56% Indels: 114  
 DB: 2 Gaps: 16

US-09-806-276a-4 (1-2546) x T30826 (1-2187)

QY 246 CCTCAGAGAGTTCTTCCATTCATTCGAAAGCCATTCCTGAAGCCACTTGAG 305  
 DB 1295 ProGluIuThrSerThrProSerProGluLysIleProLysValAlaGlyProLys 1314  
 QY 306 AAAATGATGTGACAGCTTCATCAAAAAGATTCAGAAACATATACATCTGTGAAG 365  
 DB 1315 GluIaSerAlaThrProSerLysLysThrProLysThrAlaValProLysGluThr 1334

QY 366 AGTGGCCCTTTCTCCGCTTGCAAAATAGCATTTCTCAATTCGCAAAATGCCAGCAAGA 425  
 DB 1335 SerAlaProSerGluGlyValThrAlaValProLeuGluIlePro-----ProSer 1351  
 QY 426 CCCCATTTACTTAAAGAGCCCAATTAACAAGAAAGAAAGAAATTTA----- 473  
 DB 1352 ProArg-----LysAlaProLysThrAlaAlaProLysGluThrProAlaProSer 1368  
 QY 474 -----AACTGAGGACATTTGCTCTCTGATATGATCATCCCTCCG----- 515  
 DB 1369 ProGluGlyAlaThrThrAlaProValGluIleProProSerProLysGlySerLys 1388  
 QY 516 -----TTGGAGCTTCCGACACCATTCACATTCGCAAAAGAGGCGACAGCAG 563  
 DB 1389 LysAlaGlySerLysGluThrProThrProThrProSerProGluGlyValThrAlaAlaPro 1408  
 QY 564 ATGCTTTTGAGATATTTCTTTCTTCAAGGAACATACAGACTTATCTGTAAGAACAG 623  
 DB 1409 LeuGluIle-----ProIleSerLysLysThrSerLysMetAlaSerProLys 1425  
 QY 624 AGAAGACACACTGGGCGAGTTCCCTGGGCAATATGAGTTCTCCGGGCA----- 674  
 DB 1426 GluThrLeuValThrProSerSerLysLeuSerGluThrValGlyProLysGluThr 1445  
 QY 675 -----ACAGCACT 683  
 DB 1446 SerLeuGluGlyAlaThrAlaValProLeuGluIleProProSerHisLysLysAlaPro 1465  
 QY 684 CGGACTCTGTCTCAGAAAGCCCT-----CCCGGTGCTCAAAATGCCA----- 731  
 DB 1466 LysThrValAlaProLysGluValAlaProLeuThrProSerProLysAspAlaProThrThr 1485  
 QY 732 -----TCTCCCTCCGACCAATGGAGATCCCAAGCTCATGTTCCCTATTTCT 782  
 DB 1486 LeuAlaGluSerProSerSerProLysLysAlaPro-----ThrAlaAlaProProSer 1504  
 QY 783 CACAGTGACATTTATTCACAAAGAGCTCTTGGGCGCAAGAGCTCCAGCTTA 842  
 DB 1505 GluArgValThrThrValProProGluLysProAlaThrProGluLysAlaSerGlyThr 1524  
 QY 843 GCTGCGAGCCCGTCATGAGAGAAAGAGCTCAGAGAAAGAGCTGTTGGACAAATGGGA 902  
 DB 1525 ThrAlaSer-----LysValProValProAlaGluThrGluGluValAla 1539  
 QY 903 CAGTCCACAGGAGACACT----- 923  
 DB 1540 ValSerSerArgGluThrProValThrProAlaValProProValLysAsnProSerSer 1559  
 QY 924 -----CGTGGGCTCCAGCGGTTCTGCATCTCAGT 953  
 DB 1560 HisLysLysThrSerLysThrIleGluLeuLysGluAlaProAlaThrLeuProProSer 1579  
 QY 954 CCAGGCAAGCA-----GAGACAGCCACTCTCCA 983  
 DB 1580 ProThrLysSerProLysIleProSerSerLysLysAlaProArgThrSerAlaProLys 1599  
 QY 984 GCGTTCGAGACAGTACCCGAGCTGGCCAGCCAGGACA-----TGTTGAACATC 1034  
 DB 1600 GluPhePro-----AlaSerProSerIleLysProValThrThrSerLeuAlaGluThrAla 1618  
 QY 1035 CCAAGCCATGAGAGCTCATCAAGGAAAGACTAGTCAAGAGAGCTCTCTGACCTTA 1094  
 DB 1619 ProProSerLeu-----GluLysAlaProSerThrThrIle 1630  
 QY 1095 CAGTTCCTCTCTCTCTCTGAGCTTATCTTGGGCGCTCATTGTGATGAGTGCTGA 1154  
 DB 1631 ProLysGluLysLeuAlaAlaProAlaValAla----- 1641  
 QY 1155 ATGTAATGATAAATAAGTATACAGATGCA 1187  
 DB 1642 -----ProValSerLysSerPro 1648

## RESULT 15



WMHUE2  
 HIV-EP2 enhancer-binding protein - human  
 N:Alternate names: finger protein, 275K; human immunodeficiency virus enhancer-binding P  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Mar-1993 #sequence, revision 07-Jul-1995 #text, change 21-Jul-2000  
 C:Accession: S26661; A39829; A38253  
 R:Proc. Natl. Acad. Sci. U.S.A. 89, 8971-8975, 1992  
 A:Title: Structure and expression of major histocompatibility complex-binding protein 2,  
 A:Reference number: S26661; MUID:93028387; PMID:1409593  
 A:Accession: S26661  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-2500 <VAN>  
 A:Cross-references: EMBL:X65644; NID:938259; PIDN:CAA6596.1; PID:938260  
 J. Biol. Chem. 266, 8590-8594, 1991  
 A:Title: HIV-EP2, a new member of the gene family encoding the human immunodeficiency v  
 A:Reference number: A39829; MUID:91217105; PMID:2022670  
 A:Accession: A39829  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 668-2144, R', 2146-2500 <NOM>  
 A:Cross-references: GB:M60119; NID:92661140; PIDN:AAB88218.1; PID:91822120  
 R:Rustgi, A. K.; Van't Veer, L. J.; Bernards, R.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 8707-8710, 1990  
 A:Title: Two genes encode factors with NF-kappaB- and H2TF1-like DNA-binding properties.  
 A:Reference number: A38253; MUID:91062349; PMID:2247438  
 A:Accession: A38253  
 A:Molecule type: mRNA  
 A:Residues: 1851-1990 <RUS>  
 A:Cross-references: GB:M61744; GB:M33920; NID:9187404; PIDN:AAA36202.1; PID:9187405  
 C:Genetics:  
 A:Gene: GDB:HIVEP2  
 A:Cross-references: GDB:129086; OMIM:143054  
 A:Map position: 6q23-6q24  
 C:Superfamily: HIV-EP2 enhancer-binding protein  
 C:Keywords: DNA binding; duplication; metal binding; transcription regulation; zinc fing  
 F:242-298/Region: DNA binding #status predicted  
 F:245-265/Region: zinc finger CCHH motif  
 F:273-295/Region: zinc finger CCHH motif  
 F:991-997/Region: nuclear location signal  
 F:1004-1036/Region: serine-rich  
 F:1852-1908/Region: DNA binding #status predicted  
 F:1855-1875/Region: zinc finger CCHH motif  
 F:1883-1905/Region: zinc finger CCHH motif  
 F:1953-1977/Region: acidic

Alignment Scores:  
 Pred. No.: 0 418 Length: 2500  
 Score: 115.00 Matches: 82  
 Percent Similarity: 36.88% Conservative: 36  
 Best Local Similarity: 25.62% Mismatches: 96  
 Query Match: 2.56% Indels: 106  
 DB: 1 Gaps: 14

US-09-806-276a-4 (1-2546) x WMHUE2 (1-2500)

QY 334 AGGATTCAGAAACATATCCATCTGTGAAGAAAGTGGCCCTT----- 375  
 |||||  
 Db 1391 ArgileGlnrHisValProSerTyrGlySerValMetTyrTrnSerIleSerGlnIle 1410  
 QY 376 -----TCTCCCGCT-----TGCAAAATAGACATTCATCAATTCCAAAAT 414  
 |||||  
 Db 1411 LeuGlyGlnAsnSerProAlaIleValIleCysValAsp-----GlnAsn 1426  
 QY 415 GCCAGCCAGACCAATTTTACCTGAAGCCCAATTAACAAGGAAGGAATTTAA 474  
 |||||  
 Db 1427 MetTrnGln-ArgTrnLeuValTrnAsnAlaIle-----MetGlnGlyIleGlyPheAs 1444  
 QY 475 ACTGAGGACATTCCTCTCCATGATATGATCATGTCGCCGCTTGAGACATTTGGCCACAC 534  
 |||||  
 Db 1444 nleAlaGlnValIle----- 1449

QY 535 CATCCACATGGCAAGAGGCGCAGCAGATGCTTTGGAGATATTTCTTTTCAGG 594  
 |||||  
 Db 1450 -----GlyGlnHis-AlaGlyLeuGlnLysTrpProIleTrpLysA 1463  
 QY 595 GAACACAGACCTTTTACCTGGAACACAGAGAAACACACCTGGCCAGTCCCTGGGCA 654  
 |||||  
 Db 1463 laProGlnTrnLeu-----ProLeuGlyL 1471  
 QY 655 TAATGACTTCTTCGGG-----CCACAGCAGCTCGACTGCTGTTCACAGAAC 705  
 |||||  
 Db 1471 euGlnSerSerIleProLeuCysLeuPro--SerTrnSerAspSerVal----- 1486  
 QY 706 GCCCTCCCGGTCTCAAAAATGCCATCTCTCCGACCATTTGAGAGATCCCAAGCTCT 765  
 |||||  
 Db 1487 -----AlaTrnLeuGlyGlySerLysArgMe 1495  
 QY 766 CATGTTGCCCTTATGTCTACACAGTACATTT---AATTCAAAACGAGACTCTGGGCG 822  
 |||||  
 Db 1495 TrnSerProAlaSerSerLeuGlnLeuPheMetGluTrnLysGlnLys-----Ar 1513  
 QY 823 AGCAAGCTGCCAGGCTTAGCTGCGAGCCCTCATGAGAGAAAGCTCAGCAG----- 877  
 |||||  
 Db 1513 gValLysGlnGlnLysMetLysGlnIleValGlnGlnLeuSerAlaValGlnLeuTrn 1533  
 QY 878 -----AAAGCAGTCTGTGTGAGAGATG 900  
 Db 1533 rAsnSerAspIleLysLysAspLeuSerArgProGlnLysProGlnLeuValArgGlnI 1553  
 QY 901 GACAGTCCACAG-----GGAGACACCTCGTGGGCTCCAGCGTTCATCTCA 951  
 |||||  
 Db 1553 yCysAlaSerIleProLysAspGlyLeuGlnSerGlySerSerSerPheSerLeuSe 1573  
 QY 952 GTCCAGCCAGAGCAGACAGACACCTCTCCAGCTCTCCGACAGTACCCCGACTGGCC 1011  
 |||||  
 Db 1573 rProSerSerSerGlnAspTrpTrProSerValSerProSerSerArgGlnProPhePro 1593  
 QY 1012 AGCCGAGACATGTTTGC-----CATCCCAACCCCATGAGAGCTCATCAAGGAAA 1062  
 |||||  
 Db 1593 oSerLysGlnMetLeuSerGlySerArgAlaProLeuProGlyGlnLysSerSerGlyPr 1613  
 QY 1063 GACTAAGTCAGAGAGAGTCC-----CTCTGTACCTTAC 1095  
 |||||  
 Db 1613 oSerGlnSerLysGlnSerSerAspGlnLeuAspIleAspGlnTrnAlaSerAspMetSe 1633  
 QY 1096 AGGTTCCCTCTCTCCCTGACAGCTTGATCTGGCCCTCAGCTTTGGATGAG 1147  
 |||||  
 Db 1633 rMetSerProGlnSerSerSerLeuProAlaGlyAspGlyGlnLeuGlnGln 1650

Search completed: July 18, 2003, 17:39:10  
 Job time: 121.932 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## SUMMARIES

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 18, 2003, 16:43:27 ; Search time 123.381 Seconds

(without alignments)

5499.331 Million cell updates/sec

Title: US-09-806-276A-4

Perfect score: 4500

Sequence: 1 cggggctagcccgagagacc.....cagctanaacgagggcagta 2546

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p model -DEV=xlp  
-O=/c992\_1/USPTO.spool/US09806276/runat\_18072003.160923\_12734/app\_query.fasta\_1.3854  
-DB-A=geneseq.101002 -OPMT=fstest -SUFFIX=rag -MINMATCH=0.1 -LOOPCT=0  
-LIST=45 -DUALIGN=200 -START=1 -END=1 -MATRIX=blonum2 -TRANS=human40.cdi  
-MODE=LOCAL -OUTPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US09806276.ecgn\_1.1.256 -runat\_18072003.160923\_12734 -NCP=6 -ICP=3  
-DEV\_TIMEOUT=120 -NARN\_TIMEOUT=30 -THREADS=1 -XCAPEXT=10 -LONGLOG  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A.Geneseq.101002: \*  
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: \*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: \*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: \*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: \*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: \*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: \*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: \*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: \*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: \*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: \*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: \*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: \*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: \*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: \*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: \*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: \*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: \*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: \*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: \*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: \*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: \*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: \*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Match	Length	DB	ID	Description
1	1326	29.5	254	21	AAV92240	Human bone marrow-
2	562.5	12.5	538	22	ABG25106	Novel human diagno
3	256.5	3.7	137	22	ABR27392	Human peptide #43
4	256.5	5.7	137	22	ABR32540	Peptide #46 encode
5	256.5	5.7	137	22	AAV53373	Human brain expres
6	256.5	5.7	137	22	AAV13611	Peptide #45 encode
7	256.5	5.7	137	22	AAV26008	Peptide #43 encode
8	256.5	5.7	137	22	AAV01361	Human peptide enco
9	256.5	5.7	137	22	AAV01361	Human peptide enco
10	222.5	5.0	349	22	AAE01862	Mouse zmsel protei
11	222.5	4.9	356	22	AAE03984	Human stomach can
12	222.5	4.9	356	22	AAE01861	Human zmsel protei
13	222.5	4.9	356	22	AAE03988	Human protein sequ
14	148	3.3	149	22	AAV73979	Human colon cancer
15	146	3.2	149	22	ABG03460	Novel human diagno
16	130.5	2.9	1274	22	ABE60313	Drosophila melanog
17	130.5	2.9	1356	22	ABE67291	Novel human diagno
18	129.5	2.9	819	22	ABG35569	Novel human diagno
19	129.5	2.9	1040	22	ABG14734	Novel human diagno
20	123.5	2.7	349	22	ABG14393	Novel human diagno
21	123.5	2.7	1485	21	AAV69166	A mature human N-a
22	123	2.7	697	22	AAO09987	Human polypeptide
23	121	2.7	598	22	ABG14000	Novel human diagno
24	118.5	2.6	317	21	AAV08640	Arabidopsis thalia
25	118.5	2.6	333	21	AAV08640	Arabidopsis thalia
26	118.5	2.6	333	21	AAV08640	Arabidopsis thalia
27	117	2.6	1008	22	ABE11527	Herbicidally activ
28	117	2.6	1013	22	ABE08112	Human apolipoprote
29	117	2.6	1245	18	AAV01415	Novel human diagno
30	116.5	2.6	1200	22	ABE13415	Active Factor VIII
31	116	2.6	632	22	ABE13670	Drosophila melanog
32	115.5	2.6	1331	22	ABE13670	Novel human diagno
33	115	2.6	996	23	AAE22760	Human polypeptide
34	115	2.6	1002	23	AAE22760	Human polypeptide
35	115	2.6	1475	22	AAV40232	Human PGC-3a prote
36	114.5	2.5	1015	23	AAU09420	Human PGC-3b prote
37	114.5	2.5	1574	22	ABG24228	Human PGC-3b prote
38	114	2.5	440	22	ABG05146	Mouse ECSM4 protei
39	114	2.5	539	21	AAV33050	Novel human diagno
40	114	2.5	586	20	AAV22498	Human secreted pro
41	114	2.5	586	22	AAU39067	Human secreted pro
42	114	2.5	586	23	ABE55776	Human polypeptide
43	114	2.5	884	19	AAV59666	Human acid sequenc
44	114	2.5	884	20	AAV4074	Human G-protein co
45	114	2.5	2347	18	AAV11402	Active Factor VIII

## ALIGNMENTS

RESULT 1  
AAV92240

ID AAV92240 standard; Protein: 254 AA.

AC AAV92240;

DT 10-AUG-2000 (first entry)

DE Human bone marrow-derived serum protein 2.

XX Bone marrow-derived serum protein: BMS2-2; MSB55: cytosolic; anti-HIV;

XX antiatherosclerotic; anti-inflammatory; antithrombotic;

XX antiatherosclerotic; hypotensive; antiviral; antiparasitic.

OS Homo sapiens.

XX Key

XX Region

FT

Location/Qualifiers  
1..218  
/note="homologous with MSB55 N-terminal region"

QY	413	ATGCCAGCCAGACCCCAATTACTTACCTGGAAGACGCCAATTAAACAAGAAAGAAAGAAATTT	472
Db	1	MeerProAlaIalysThrProIleTyrLeuIysAlaAlaAsnLysLysGlyLysLysPhe	20
QY	473	AAACGTGAGGACATTTGTGTCTCTCTATATGATGACGTCCCGCTTGAGACTTTGGCCAC	532
Db	21	LysLeuIarGspIleLeuSerProAspMetIleSerProPoleuGlyAspPheArgHis	40
QY	533	ACCAATCCACATTTGGCAAGAAGGGCCAGCACATGTCTTTGGAGATATTTCCTTCTTCAA	592
Db	41	ThrIleHisTleGlyLysGlnGlyGlnHisAspValPheGlyAspIleSerPheLeuGln	60
QY	593	GGGAACTACGACGCTTTTACTGGAAACCCAGGAGAAAGACACACCTGGGCCAGTTTCCCTGGG	652
Db	61	GlyAsnTyrGlnLeuLeuProGlyAsnGlnGlnLysAlaHisIleuGlyIlePheProGly	80
QY	653	CATATATGAGTTCTTCCGGGCCAACAGCACCTGGACTCTGTGTTCACAAAGAGCCCTCC	712
Db	81	HisAsnGlnPhePheArgAlaAsnSerThrSerAspSerValPheThrGlnIleuProSer	100
QY	713	CCGGTGTCCAAAATGCCATTCCTCCCTCCGACCTTGAGAGATCCCAAGCTCTCATGTTGG	772
Db	101	ProValIleuLysAsnAlaIleSerLeuProThrIleGlyLysArgIleAlaIleuMetLeu	120
QY	773	CCCTATTGTCCACAGTACATTTAATTCACAACAGACATCTTCCGGGCCAGCAAAAGCTG	832
Db	121	ProIleuLeuSerProValThrPheAsnSerLysGlnIleuSerPheGlyProAlaLysLeu	140
QY	833	CCCAAGCTTACGTCGGAGCCCGCTCATGGAGGAAAAGCTCAGAGAAAGACAGTCTGTTGG	892
Db	141	ProArgIleuSerCysGlnProValMetGlnLysAlaGlnGlnLysSerSerLeuLeu	160
QY	893	GAGAAATGGGACAGTCCACCAGGAGAGACACCTCTGGGGCTCCAGCGCTTCGCATTCACG	952
Db	161	GlnAsnGlyThrValHisGlnGlyAspThrSerTrpGlySerSerGlySerAlaSerGln	180
QY	953	TCCAGCCAAAGCAGAGACAGCCACCTCCCTCCACCTGTCGGAACAGTACCCGACTGGCCA	1012
Db	181	SerSerIleGlnIlyArgAspSerHisSerSerLeuSerGlnGlnIlyProAspTrpPro	200
QY	1013	GCCAGAGACATGTTTGACATCCACCCCATGGAGGAGCTCATCAAGGGAACAGTAAAGCA	1072
Db	201	AlaGlnLysPheIleAspHisProIleProCysGlnLeuIleLysGlyLysThrLysSer	220
QY	1073	GAGAGATCCCTCTGTACCTTACAGGTTCCCTCTCCCTCCAGCTTGATCTTGGGCC	1132
Db	221	GlnIleuSerLeuSerAspLeuIleGlySerLeuLeuSerLeuGlnIleuAspLeuPro	240
QY	1133	TCACCTTTGGATGAGTGTGTAATGATGATGAATAAATTAAG	1174
Db	241	SerLeuLeuAspGlnValLeuAsnValMetAspLysAsnLys	254
RESULT 2			
ABG25106			
ID	ABG25106	standard; Protein: 538 AA.	
AC	ABG25106:		
XX	18-FEB-2002	(first entry)	
DE	Novel human diagnostic protein #25097.		
XX	Human: chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder.		
OS	Homo sapiens.		
XX	W0200175067-A2.		
XX	11-OCT-2001.		
XX	30-MAR-2001; 2001WO-US08631.		
PF			
XX			



CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

Sequence 137 AA;

Alignment Scores:

Pred. No.:	1,296-15	length:	137
Score:	256.50	Matches:	56
Percent Similarity:	63.46%	Conservative:	10
Best Local Similarity:	53.85%	Mismatches:	35
Query Match:	5.70%	Indels:	3
DB:	22	Gaps:	2

US-09-806-276A-4 (1-2546) x ABB27392 (1-137)

[illegible]

## RESULT 4

ID ABB32540 standard; Peptide; 137 AA.

AC	ABB32540;
XX	
DT	01-FEB-2002 (first entry)

Peptide #46 encoded by human foetal liver single exon nucleic acid probe.

KW Human; foetal liver; gene expression; single exon nucleic acid probe.

OS Homo sapiens.

PN W0200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US006669.

AA		2000US-0180312
PR	04-FEB-2000	2000US-0207456
PR	26-MAY-2000	2000US-0608408
PR	30-JUN-2000	2000US-0652366
PR	03-AUG-2000	2000US-0234687
PR	21-SEP-2000	2000US-0236359
PR	27-SEP-2000	2000GB-0024265
PR	04-OCT-2000	

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver -

PS Claim 27; SEQ ID NO 25175; 639pp + sequence listing; English.

CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at [http://wipo.int/pub/published\\_pat\\_sequences](http://wipo.int/pub/published_pat_sequences).

50 Sequence 137 AA;

Alignment Scores:	
Pred. No.:	1,39e-15
Score:	256.50
Percent Similarity:	63.46%
Best Local Similarity:	53.85%
Query Match:	5.70%
DB:	22
Length:	137
Matches:	56
Conservative:	10
Mismatches:	35
Indels:	
Gaps:	2

US-09-806-276A-4 (1-2546) X ABB32540 (1-137)

OY	411	ATGGCAGCCAGACCCCAATTACCTGAAGAAGCAGCCATTAAACAAGAAAGAAACAATTT	472
Db	35	MelSerThrIlysalProIletryIleuys---ArgIlySerAlrgIlyLysIyLysGlu	53
OY	473	AAACTGAGGACATCTGTCTCCTGATATGATGACGTCCCGCGCTGGAGACTTTGCCAC	532
Db	54	LysIleuArIgrsPneuIeuSerIserAspMetIieserProIleuGIyAspPheArgHis	73
OY	533	ACCATCCACATWTGGCAAAAGGCGCCAGCATCTGTTGGAGATATTTCCPTCTTCAA	592
Db	74	ThrlIeIstIleIyIserGIyGIyGIyIserAspMetPheGIyAspIleIserPheIeuGln	93
OY	593	GGGAACTACGAGCTTTTACCCTGGAAACAGGAAAGACA-----CACCTGGGCCAAGTTC	646
Db	94	GIyLysPheHisIleuIeuProGIyThrMetValGIuGIyProGIuIuAspGIyThrPhe	113
OY	647	CTGTGGCATATATAGTCTTTCGGGGCAACAGCAGCACTGGAGCTGTGTTACAGAAACG	706
Db	114	AspIeuProPheGlnPheThrAlrArgThralaIthrValCysGIyArGIuIeuProAspGIy	133
OY	707	CCCTCCCCGGTG	718
Db	134	ProIserProIleu	137

## RESULT 5

ID AAM53373 standard; Protein; 137 AA.

AC 22M53373;

DT 05-NOV-2001 (first entry)

Human brain expressed single exon probe encoded protein SEQ ID NO: 25478.

Human; brain expressed exon; gene expression analysis; probe; microarray; Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.

OS Homo sapiens

PN W0200157275-A2.

09-AUG-2001  
PD

30-JAN-2001: 2001WO-US006667.

04-FEB-2000: 2000US-0180312  
XX  
PR







Percent Similarity:	63.46%	Conservative:	10
Best Local Similarity:	53.85%	Mismatches:	35
Query Match:	5.70%	Indels:	3
DB:	22	Gaps:	2

US-09-806-276A-4 (1-2546) x AAM01361 (1-137)

[illegible]

RESULT 9  
ABG35381  
ID ABG35381 standard; Peptide; 137 AA

AC	ABG35381;
XX	
DT	19-AUG-2002 (first entry)
XX	
DE	Human peptide encoded by genome-derived single exon probe SEQ ID 25046

KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW Chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagazer syndrome  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesias; pulmonary hypertension;  
 KW hyaline membrane disease.

OS Homo sapiens.

PN WO200186003-A2

PD 15-NOV-2001

PF 30-JAN-2001; 2001WO-US00665

PR 04-FEB-2000; 2000US-180312P  
PR 26-MAY-2000; 2000US-207456P  
PR 30-JUN-2000; 2000US-060840S  
PR 03-AUG-2000; 2000US-0632366  
PR 27-SEP-2000; 2000US-234687P  
PR 21-SEP-2000; 2000US-236359P  
PR 04-OCT-2000; 2000GB-0024263

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR

DR WPT; 2002-114183/15

PT Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 XX  
 PS Claim 27; SEQ ID No 25046; 634pp; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12872 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes: the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression in a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karsagen syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published\\_pcl\\_sequences](http://wipo.int/pub/published_pcl_sequences).

Sequence 137 AA.

Alignment Scores:	
Pred. No.:	1,29e-15
Score:	256.50
Percent Similarity:	63.46%
Best Local Similarity:	53.85%
Query Match:	7.0%
DB:	23
	Gaps: 2

US-09-806-276A-4 (1-2546) x ABG35381 (1-137

```

QY 412 ATGCAGCCAGAACCCCAATTATTCGTGGAAGACGCCATTAAACAGAAAGAAAGAAATTT 472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35 MetSerThrIysValProIleTyrLeuLys---AArgGlySerArgLysGlyLysLysGlu 53

QY 473 AACTGAGGAGCAATTCGTCTCCTGATATGATGACAGTCCCGCTTGAGACTTTGGCCAC 532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54 LysLeuArgPheLeuLeuSerSerAspMetIleSerProProLeuGlyAspPheArgHis 73

QY 533 ACCATCCCAATTGGCAAGAAGGCCACAGATGCTTTGGAGATATTTCTTTCTTCAA 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 74 ThrIleHisIleGlySerGlyGlySerAspMetPheGlyAspIleSerPheLeuGln 93

QY 593 GGGAACTACAGACTTTTAACTGTGAACACAGAGAAAGCA-----CACTGTGGGCAAGTC 646
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 GlyLysPheHisLeuLeuProGlyIleThrMetValGlnGlyProGlnGlyAspGlyThrPhe 113

```

QY 647 CCTGGCATTAATGACTTCTTCGGGCGCAACAGACCTCGACTCTGTGTTACAGAAACG 706  
 Db 114 ASpleuProPheGlnPheThrAlaThrValCysGlyArgGluLeuProAspGly 133  
 QY 707 CCCTCCCGCGG 718  
 Db 134 ProSerProLeu 137  
 RESULT 10  
 AAE01862  
 ID AAE01862 standard; Protein; 349 AA.  
 AC AAE01862;  
 XX  
 DT 31-JUL-2001 (first entry)  
 XX  
 DE Mouse zmsel protein.  
 XX  
 KW Mouse; zmsel protein; Cdc42/Rac interactive binding protein; CRIB;  
 KW Wiskott-Aldrich Syndrome; cancer; tumour; invasion; metastasis; asthma;  
 KW digestion; actin polymerisation; cytoskeletal reorganisation; arthritis;  
 KW testicular function; muscle inflammation; inflammatory bowel disease;  
 KW diverticulitis; male infertility; male contraceptive agent; myocarditis;  
 KW spermatogenesis; sperm capacitation; reperfusion ischaemia; psoriasis;  
 KW melanoma; atherosclerosis; pelvic inflammatory disease; PID; eczema;  
 KW scleroderma; cytoskeletal; vasotropic; dermatological; gene therapy.  
 XX  
 OS Mus musculus.  
 FH Key  
 FH Location/Qualifiers  
 FT 1..145  
 FT /note= "Conserved N-terminal domain"  
 FT Domain  
 FT 27..41  
 FT /label= CRIB\_motif  
 FT 146..329  
 FT /note= "Variable C-terminal domain"  
 FT 329..350  
 FT Region  
 FT /note= "Highly conserved C-terminal tail"  
 XX  
 PN WO200134803-A2.  
 PD 17-MAY-2001.  
 XX  
 PE 09-NOV-2000; 2000WO-US30945.  
 PR 10-NOV-1999; 99US-0438564.  
 XX  
 PA (ZYMO ) ZYMOGENETICS INC.  
 XX  
 PI Holloway JL, Gao Z, Whitmore TE;  
 XX  
 DR WPI: 2001-335928/35.  
 DR N-PSDB: AAD05673.  
 XX  
 PT Novel human CRIB protein, zmsel and polynucleotide encoding the  
 PT protein, for detecting human chromosomal abnormalities and for treating  
 PT cancer, cardiovascular and inflammatory conditions  
 XX  
 PS Claim 21; Fig 2; 132pp; English.  
 XX  
 CC The present invention relates to DNA and protein for zmsel, a novel  
 CC human Cdc42/Rac interactive binding (CRIB) protein. CRIB proteins are  
 CC implicated in human disease such as Wiskott-Aldrich Syndrome. zmsel  
 CC modulators are useful for modulating tumour cell motility, invasion and  
 CC metastasis, gene transcription, contractility of various tissues, actin  
 CC polymerisation and cytoskeletal reorganisation, digestion, testicular  
 CC function and fertility. zmsel sequence and its modulators are useful for  
 CC treating cancer, inflammatory heart or cardiovascular conditions, muscle  
 CC inflammation, inflammation during and after surgery, arthritis, asthma,  
 CC inflammatory bowel diseases or diverticulitis, myocarditis, scleroderma,  
 CC atherosclerosis, pelvic inflammatory disease (PID), eczema and other  
 CC inflammatory diseases, male infertility or as male contraceptive agents

CC and for modulating spermatogenesis and sperm capacitation. zmsel and  
 CC anti-zmsel antibodies are useful in diagnosing inflammatory diseases,  
 CC such as reperfusion ischaemia, psoriasis, arthritis, melanoma and other  
 CC inflammatory diseases, male reproductive cancers such as prostate and  
 CC testicular cancers. zmsel polynucleotide sequences are useful as probes  
 CC or primers for detecting human chromosomal abnormalities. zmsel sequence  
 CC is used in gene therapy. The present amino acid sequence is mouse  
 CC zmsel protein.  
 XX  
 SQ Sequence 349 AA;  
 Alignment Scores:  
 Pred. No.: 3,23e-12 Length: 349  
 Score: 223.00 Matches: 76  
 Percent Similarity: 44.578 Conservative: 43  
 Best Local Similarity: 28.466 Mismatches: 82  
 Query Match: 4.968 Indels: 66  
 DB: 22 Gaps: 10  
 US-09-806-276a-4 (1-2546) x AAE01862 (1-349)  
 QY 428 CCAATTACCTGAAAGCAGCCATTAAGAAAGAAATTTAACTGAGGACATT 487  
 Db 2 ProileLeuylGlnLeuValSerSerValAsnSerLysArgArgSerAlaAsp 21  
 QY 488 CTGTCTCTGATATGATCATGTCCTCCCGCTTGAGACTTTGGCCACCATCCACTTGGC 547  
 Db 22 LeuThrAlaGlnMetIleSerAlaProLeuGlnLysAspPheArgHisThrMetHisValGly 41  
 QY 548 AAAGAGCGCCACAGCATGCTTTGGAGATATTCTCTCTT-----CAA 592  
 Db 42 ArgAlaGly-----AspAlaPheGlnLysPheThrSerPheLeuThrSerLysAlaArgGlu 59  
 QY 593 GGGACTACGACTTTTACCTGTAAGCAAGAAAGCAACCTGGCCAGTTCCTGGG 652  
 Db 60 AlaAspAspGlnSerLeu-----AspGlnGlnAlaSerAlaSerLysLeuSerLeu 76  
 QY 653 CATTAATGACTTCTTCGGGCGCAACAGCACTGGCACTGTG----- 694  
 Db 77 LeuSerArgLysPheArgLysSerLysArgSerGlnSerValThrArgGlyAspArgGlu 96  
 QY 695 -----TTCAAGAAAGCGCCCTCCCGGTCTCAAAATGCGCATCTCC 736  
 Db 97 GlnArgAspMetLeuGlnSerLeuArgAspSerAlaLeuPheValLysAsnAlaMetSer 116  
 QY 737 CTCGCGACCATTTGGAGATCCCAAGCTCTCATGTTGCCCTTATTGTACACAGTACATT 796  
 Db 117 LeuProGlnLeuAsnGlnLysGlnAla----- 125  
 QY 797 AATTCCAACAGAGAGTCTTCGGGCGCCAGCAAGCTGCCAG-----CTTAGCTGGAGGCC 853  
 Db 126 -----AlaGlnLysAspSer-----SerLysLeuProLysSerLeuSerPro 141  
 QY 854 GTCATGAGAGAAAGAACTCAGAGAGAAAGAGTCTGTGGAGATGGACAGTCCACAG 913  
 Db 142 ValLysLysAlaAspAlaArgAsp----- 149  
 QY 914 GGAAGACCTCTGGGGCTCCAGCGGTTCTGCATCTCAGTCCAGCAAGCAGACAGC 973  
 Db 150 -----GlyGlyProLysSerProHisArgAsnGlnLysAlaThrGlyProAsnSer 165  
 QY 974 CACTCTCAGCCTTCGAGACAGATACCCGACGCGGCGAGCGGAGCATGTTGGACAT 1033  
 Db 166 ProAspProLeuLeuAspGln-----AlaPheGlnLysPheLeuMetAspLeu 181  
 QY 1034 CCCACCCATGCGAGCTCATCAAGGAAAGAGTCAAGAGAGTCCCTCTTGACCTT 1093  
 Db 182 ProIleMetProLysValSerTyrGlyLeuLysHisAlaGlu----- 195  
 QY 1094 ACAGTTCCT 1153  
 Db 196 -----SerIleLeuSerPheHisIleAspLeuGlyProSerMetLeuGlyAspValLeu 213



FT Domain  
 FT /note= "Conserved N-terminal domain"  
 FT 14..147  
 FT Modified-site  
 FT /note= "Phosphorylation site"  
 FT 14..16  
 FT Domain  
 FT /label= CRIB\_motif  
 FT 27..41  
 FT Modified-site  
 FT /note= "Phosphorylation site"  
 FT 80..82  
 FT Modified-site  
 FT /note= "Phosphorylation site"  
 FT 86..88  
 FT Region  
 FT /note= "Hydrophilic region"  
 FT 96..101  
 FT Modified-site  
 FT /note= "Phosphorylation site"  
 FT 105..107  
 FT Domain  
 FT /note= "Variable C-terminal domain"  
 FT 148..336  
 FT Region  
 FT /note= "Hydrophilic region"  
 FT 226..231  
 FT Modified-site  
 FT /note= "Phosphorylation site"  
 FT 295..297  
 FT Modified-site  
 FT /note= "Phosphorylation site"  
 FT 303..305  
 FT Region  
 FT /note= "Highly conserved C-terminal tail"  
 FT 337..356  
 FT Region  
 FT /note= "Hydrophilic region"  
 FT 346..351  
 FT Region  
 FT /note= "Hydrophilic region"  
 FT 347..352  
 FT Region  
 FT /note= "Hydrophilic region"  
 FT 348..353  
 FT Region  
 FT /note= "Hydrophilic region"  
 FT  
 PN WO200134803-A2.  
 PD 17-MAY-2001.  
 XX 09-NOV-2000; 2000MO-US30945.  
 XX 10-NOV-1999; 99US-0438564.  
 XX (ZYMO ) ZYMOGENETICS INC.  
 PA (ZYMO ) ZYMOGENETICS INC.  
 XX  
 PI Holloway JL, Gao Z, Whitmore TE;  
 XX WPI: 2001-335928/35.  
 DR N-PSDB: AAD05671.  
 XX  
 PT Novel human CRIB protein, zmsel and polynucleotide encoding the  
 PT protein, for detecting human chromosomal abnormalities and for treating  
 PT cancer, cardiovascular and inflammatory conditions  
 PT  
 PS Claim 11; Fig 2; 132pp: English.  
 XX  
 XX The present invention relates to DNA and protein for zmsel, a novel  
 CC human Cdc42/Rac interactive binding (CRIB) protein. CRIB proteins are  
 CC implicated in human disease such as Wiskott-Aldrich Syndrome. Zmsel  
 CC modulators are useful for modulating tumour cell motility, invasion and  
 CC metastasis, gene transcription, reorganisation, digestion, testicular  
 CC polymetastasis and cytoskeletal reorganisation, various tissues, actin  
 CC function and fertility. Zmsel sequence and its modulators are useful for  
 CC treating cancer, inflammatory heart or cardiovascular conditions, muscle  
 CC inflammation, inflammatory during and after surgery, arthritis, asthma,  
 CC inflammatory bowel diseases or diverticulitis, myocarditis, scleroderma,  
 CC atherosclerosis, pelvic inflammatory disease (PID), eczema and other  
 CC inflammatory diseases, male infertility or as male contraceptive agents  
 CC and for modulating spermatogenesis and sperm capacitation. zmsel and  
 CC anti-zmsel antibodies are useful in diagnosing inflammatory diseases,  
 CC such as reperfusion ischaemia, psoriasis, arthritis, melanoma and other  
 CC inflammatory diseases, male reproductive cancers such as prostate and  
 CC testicular cancers. Zmsel polynucleotide sequences are useful as probes  
 CC or primers for detecting human chromosomal abnormalities. zmsel sequence  
 CC is used in gene therapy. The present amino acid sequence is human  
 CC zmsel protein. The zmsel gene is located at the 17q24.1 region of  
 CC chromosome 17.

XX SQ Sequence 356 AA:  
 Alignment Scores:  
 Pred. No.: 3.65e-12 Length: 356  
 Score: 222.50 Matches: 77  
 Percent Similarity: 41.91% Conservative: 37  
 Best Local Similarity: 28.31% Mismatches: 91  
 Query Match: 4.94% Indels: 67  
 DB: 22 Gaps: 11  
 US-09-806-276a-4 (1-2546) x AAE01861 (1-356)  
 QY 428 CCAATTTACCTGGAAGCAGCCAAATTAACAGAAAGAAATTTAACTGAGGCACAT 487  
 DB 2 ProileuLysGlnLeuValSerSerValHisSerLysArgArgSerArgAlaasp 21  
 QY 488 CTGTCTCCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 547  
 DB 22 LeuThrAlaGluMetIleSerAlaProLeuGlyAspPheArgHisThrMetHisValGly 41  
 QY 548 AAGAGGGCCGACGACGATGCTTTGGAGATATTTCTTCTTCAAGGAACTACGAGCTT 607  
 DB 42 ArgAlaGly-----AspAlaPheGlyAspThrSerPheLeuAsnSerLys----- 56  
 QY 608 TTACCTGGAACAGAGAAAGACACACACCTGGCCGATTCCTCCGATTAATGAGTTC--- 664  
 DB 57 ---AlaGlyGluProAspGlyGluSerLeuAspGluInProSerSerSerSerLys 75  
 QY 665 -----TTCCGGCCCAACAGACACCTCGGACCTGTGTG----- 694  
 DB 76 ArgSerLeuLeuSerArgLysPheArgLysSerArgSerGlnSerValThrArgGly 95  
 QY 695 -----TTACAGAAAGCCCTCCCGGTCCTCAAAAT 727  
 DB 96 GluArgGluGlnArgAspMetLeuLysSerLeuArgAspSerAlaLeuPheValAsn 115  
 QY 728 GCCATCTCCCTCCGACCATTTGGAGAGATCCAGCTCATGTTGCCCTTATTGTACCA 787  
 DB 116 AlaMetSerLeuProInLeuAsnGluLysGluAla----- 127  
 QY 788 GTGACATTATTTCCAAAGAGAGCTCTGGCGCCGATTCGATTCAGTCCAGCAGC---CTTAC 844  
 DB 128 -----AlaGluLysGlyThrSerLysLeuProLysSerLeuSer 140  
 QY 845 TGCAGCCGCTCATGAGGAGAAAGCTCAGGAAAGCAAGCTGTGTGGAGAAATGGACA 904  
 DB 141 SerSerProVal-----LysLysAlaAsnAspGlyGly----- 152  
 QY 905 GTCCACCGGAGACACCTCGTGGGCTCCAGCGTTCGATTCAGTCCAGCAGCAGC 964  
 DB 153 -----GlyAspGluGluAlaGlyThrGluGluAlaValProArgArgAsnGlyAla 169  
 QY 965 AGAGACAGCCACCTCCAGC-----CTTCCGAAACAGTACCCGAGTGGCCAGCCGAG 1018  
 DB 170 AlaGlyProHisSerProAspProLeuAspGluIn-----AlaPheGly 185  
 QY 1019 GACATGTTTACCATCCACCCATCGAGCTCATCAAGGAAAGAACTAAGCAAGCAGAG 1078  
 DB 186 AspLeuThrAspLeuProValAlaProLysAlaThrTyrgLysLeuLysHisAlaLys 204  
 QY 1079 TCCTCTCTGACCTTACAGTTCCTCCCTCCCTGACCTGATCTGGGCGCTTACCT 1138  
 DB 205 -----SerIleMetSerPheHisIleAspLeuLysProSerMet 217  
 QY 1139 TTGATGAGGTGCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1174  
 DB 218 LeuGlyAspValLeuSerIleMetAspLysGluGlu 229  
 RESULT 13  
 AAB95088  
 ID AAB95088 standard; Protein: 356 AA.  
 XX



Fri Jul 25 09:59:18 2003

us-09-806-276a-4.rag

28-SEP-2000; 2000MO-US26524.  
 XX 29-SEP-1999; 99US-0157137.  
 PR 03-NOV-1999; 99US-0163280.  
 XX  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
 PI WPI: 2001-235357/24.  
 XX N-PSDB: AAH33410.  
 DR  
 DR  
 PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
 XX useful for preventing, diagnosing and/or treating colorectal cancers -  
 PS Claim 11: Page 6543-6544; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patient's own production of P.  
 CC Additionally, N may be used to produce the colon cancer-associated P,  
 CC by inserting the nucleic acids into a host cell and culturing the cell  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
 CC and AAB77789 represent sequences used in the exemplification of the  
 CC present invention.  
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

Alignment Scores:  
 Pred. NO: 3.16e-05 Length: 149  
 Score: 148.00 Matches: 61  
 Percent Similarity: 35.75% Conservative: 13  
 Best Local Similarity: 29.47% Mismatches: 46  
 Query Match: 3.29% Indels: 87  
 DB: 22 Gaps: 12

US-09-806-276a-4 (1-2546) x AAG73979 (1-149)  
 QY 395 ACATTCTCAATTCCTCAAAATGCCAGCCAGCCCAATTTAC----- 436  
 DB 1 SerTPSerArGuaTGAAGAlaProAla-----ProLeuTrpGluAspArgGluMetPro 18  
 QY 437 ---CTGAAGACGACCAATACAGAAAGAAAGAAATTTAACTGAGGAGACTTCTGCT 493  
 DB 19 ValLeuLysGlnLeuGlyProAlaGlnProLysArgProAspArgGlyAlaLeuSer 38  
 QY 494 CCTGATATGATCAGTCCCGCTTGAGACTTTCGCACACCATCCATCGCAATGGCAAGAG 553  
 DB 39 -----TLeSerAlaProLeuGlyAspPheArgHisThrLeuHisValGlyArgGly 55  
 QY 554 GGCACGACGATGCTTTTGAGATATTTCTTTCTTCAAGGAAACTGACAGCTTTTACCT 613  
 DB 56 Gly-----AspAlaPheGlyAspPheSerPheLeu----- 65  
 QY 614 GGAACCGAGAGAAACACACACTGGGCGGCTCCCTGGGCAATATGACTTCTCCGGGCC 673  
 DB 66 -----SerArgHisGlyGly-Gly----- 71  
 QY 674 AACAGCACTCGACTCTGTGTTCACAGAAAGCCCTCCCGCGTCTCAAAATGCCATC 733  
 DB 72 -----ProPro-----ProSe 75

QY 734 TCCCTCCCGACCATGTGAGAGATCCCAAGCTCTCATGTTCCTTATTTGTCACCAAGTAC 793  
 DB 75 rProGlyArgProProArgGlyProArg----- 84  
 QY 794 TTTAATTCCAACAGAGACTCTCGGCGACGCAAGACTGCCAGGCTTA---GCTGGCAG 850  
 DB 85 -----ProArgArgArgArgArgProGlnSerAlaAlaProAlaGlnArgProAlaVa 103  
 QY 851 CCCGTCA-----TGGAGGAAAGAGCTCAGAGAAAGCAGCTCG 889  
 DB 103 lProSerProGlySerGlyAlaSerCysTrpThrArg-----GACACACTCGCG 928  
 QY 890 TTGAGATGGCAGCTCCACACAG----- 116  
 DB 116 strArgMetAlaAlaAlaArgArgSerGlyCysAlaSerHisAlaAsnProProCly\*\* 136  
 QY 929 GGCTCCAGCGGTTCGCAT 947  
 DB 136 \*AlaProAlaValArgHis 142

RESULT 15  
 ABG03460  
 ID ABG03460 standard; protein: 476 AA.  
 XX  
 AC ABG03460;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #3451.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 XX  
 PN W0200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001MO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 XX  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 PI N-PSDB: AAS67647.  
 DR WPI: 2001-639362/73.  
 DR  
 DR  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PT  
 PS Claim 20: SEQ ID No 33819; 103pp; English.

The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. [AB000010-AB030377](#) represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPAC  
CC at [ftp://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
XX  
Sequence 476 AA;

Alignment Scores:	
Pred. No.:	8,6e-05
Score:	1,0e+00
Percent Similarity:	3,0e+00
Best Local Similarity:	2,7e-08
Query Match:	3,4e-06
DB:	22
	Gaps:
	21

000-2/0A-4 (1-2546) x ABG03460 (1-476)

[illegible][illegible]

Search completed: July 18, 2003, 17:24:52  
Job time : 134.381 secs





July 18, 2003, 17:19:12 ; Search time 27 9421 seconds

(without alignments)  
5361 850 4114

title: US-09-806-276A-4  
perfect score: 4500

**sequence:**

1 cggggctagcccgagacc...Gagtanaaaccaacattc acac

scoring table:

Xgapop	10.0	Xgapext	0.5
Ygapop	10.0	Ygapext	0.5
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

```

Post-processing: Minimum

Listing first 45 summaries

Command line parameters:  
-MODEL=fuzzy

```

0#model=11name=n2p.model DEV=xlp
0#Gm2=11usftp.spool/US09806276/runat.18072003.160925.12832/app-query.fasta-1.385
DB-issued.patents.AA -OEMT=fstestn -SUFFIX=ara1 MINMATCH=0.1 -LOOPCC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blissumc2 -TRANS=humano4.cdi
-LIST=45 -DOCLINK=200 -THR_SCORE=pcet -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUPM=ptc -NORM=ext HEAPSIZED=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US09806276.ecgn-1.1.56.#runat.18072003.160925.12832 -NCPu=6 -ICP=3
-NO_MAP LARKEOVER -NEG_SCORES=0 -WAIT -DSBLOK=100 -LONGLOG
DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XCAPOP=10 -XGAPEXT=0.5 -FGAPEXT=7
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database : Issued\_Patents\_AA:\*

```

1: Issued/patents_AA.*
2: /cgn2_6/prodata/1/iaa/5A_COMB.pcp.*
3: /cgn2_6/prodata/1/iaa/5B_COMB.pcp.*
4: /cgn2_6/prodata/1/iaa/6A_COMB.pcp.*
5: /cgn2_6/prodata/1/iaa/6B_COMB.pcp.*
6: /cgn2_6/prodata/1/iaa/PTCUS_COMB.pcp.*
7: /cgn2_6/prodata/1/iaa/backfiles1.pcp.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	114	2.5	884	2	US-08-465-976A-2	Sequence 2, Appl
2	114	2.5	884	2	US-08-465-976A-2	Sequence 2, Appl
3	110.5	2.5	183	6	5168049-3	Sequence 2, Appl
4	110.5	2.5	1117	2	US-08-982-412-2	Patent No. 516804
5	110.5	2.5	1117	2	US-08-841-483-6	Sequence 6, Appl
6	110	2.4	550	5	US-09-382-911-6	Sequence 6, Appl
7	110	2.4	2332	1	PCT-US93-0202-1	Sequence 1, Appl
8	110	2.4	2332	1	US-07-864-004B-4	Sequence 1, Appl
9	110	2.4	2332	1	US-08-251-937A-4	Sequence 4, Appl
10	110	2.4	2332	1	US-08-212-133A-2	Sequence 4, Appl
11	110	2.4	2332	1	US-08-474-503-2	Sequence 2, Appl
12	110	2.4	2332	1	US-08-670-707A-2	Sequence 2, Appl
			2332	4	US-09-037-601-2	Sequence 2, Appl

[illegible]

## ALIGNMENTS

RESULT 1

US-08-485-9/bA-2  
; Sequence 2, Application US/08465976  
; Patent No. 5869632

GENERAL INFORMATION

APPLICANT: I.T. VI

APPLICANT:

TITLE OF INVENTION: RUBEN, STEVEN I

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: STEWART & OLSTEIN

STREET: 6 BECKER FARM ROAD

STATE: NJ

COUNTRY: U

DATE: 07/008  
COMPUTER READ

MEDIUM TYPE

COMPOILER: IBM PC compatible

SOFTWARE: PatentIn Release

CURRENT APPLICATION DATA:

FILING DATE: 06-JUN-1995

CLASSIFICATION: 42.

NAME: FERRARO GREGORY F

REGISTRATION NUMBER: 36,134

TELECOMMUNICATION INFORMATION  
REFERENCE/DOCUMENT NUMBER: 3

TELEPHONE: (201) 994-1770

TELEFAX: (201) 994-1744

SEQUENCE CHARACTERISTICS:  
 LENGTH: 884 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-465-976a-2

## Alignment Scores:

Pred. No.: 0.00518 Length: 884  
 Score: 114.00 Matches: 93  
 Percent Similarity: 30.02% Conservative: 31  
 Best Local Similarity: 22.52% Mismatches: 125  
 Query Match: 2.53% Indels: 164  
 Gaps: 21

US-09-806-276a-4 (1-2546) x US-08-465-976a-2 (1-884)

QY 6 CTAGCCCGGAGACCCGGCCCTGGGCGCTTCACGCGCTCTCGAGCGGATTAAT 65  
 DB LeuGlyArgAlaGlnArgGlnArgGlyLeuLeuysAsnArgLeuGlnTyrProLeu 581  
 QY 66 GCGGTAGACAGGACACCGCGGACGACTCGGCTGATTCGCGACAGCGGCGGATTTG 125  
 DB 582 ValProGlnThrArgGlyAlaProGlnLeuSerTyrPysArgAlaAlaThrLeuGlyHis 601  
 QY 126 CGTGGC---CCGGGGGAGGCGCGGCGGCTGGGATCTCAGCGGCGCGGCTTTG 182  
 DB 602 ArgAlaValProAlaAlaSerTyrGlyArg-----IleTyrAlaGly--- 615  
 QY 183 TCCTGTGTGTGTCAGACGATGATTAATGCTCTCTAGAGACCTTCACCTTGGCCG 242  
 DB 616 -----GlyGlyThrGlySerLeuSerGlnPro 624  
 QY 243 TAACTCAGAGAGTCT-----CTTGAACCCATCTCATTTGAGAG 284  
 DB 625 AlaSerArgTyrSerArgGlnLeuAspLeuLeuArgArgGlnLeuSerArg 644  
 QY 285 -----CCACTTCTGAAGCCACTTGAAGAAAATGATGTGAC 320  
 DB 645 GlnArgLeuGlnGlnAlaProAlaProValLeuArgProLeuSerArg----- 660  
 QY 321 GTTCCTATCAAAAAGATTCAAGACATATACCATCTGAGAGAAAGTGGCCCTTCTCC 380  
 DB 661 ---ProGlySerGlnGlu-----CysMetAspAlaAlaPro---Gly 672  
 QY 381 CGCTTGCAG-----AATAGACATCTCAATTCAGAAATGCCAGCCAGACCCCAATT 434  
 DB 673 ArgLeuGlnProLysAspArgGlySerThrLeuProArgArgGlnProProArgAspTyr 692  
 QY 435 ACCTGAAGACGACCAATTAAGAGAAAGAAATTTAACTGAGGACATTTCTGTCTC 494  
 DB 692 ----- 692  
 QY 495 CTGATATGATGATGATCCCGCTGGAGACTTTCGCACACCATTCATTTGGCAAGAG 554  
 DB 693 -----ProGlyAlaMetAlaGlyArg 699  
 QY 693 ----- 699  
 QY 555 GCCAGCAGCATGCTTTGGAGATATTTCTTTCTCAAGGACATCAGAGCTTTTACTGTG 614  
 DB 700 PheGlySerArgAsp----- 704  
 QY 615 GAACACGAGAGAAAGACACCTGGGCAAGTCCCTGGGCAATATGAGATTCTTCCGGGCCA 674  
 DB 705 -----AlaLeuAspLeuGly----- 709  
 QY 675 ACAGCAGCTCGACTCTGTTCACAGAAAGCCCTCCCGGTGCTCAAAAATGTCATCT 734  
 DB 710 ---AlaProArgGlnTyrLeuSerThrLeuProProProArg----- 722  
 QY 735 CCCTCCCGCAGCATTTGGAG---GATCCCAAGCTCTCATGTTGCCCTTATTTGTACACAGTGA 791  
 DB 723 ArgThrArgAspLeuAspHisSerProHisLeu-----CysPro----- 735

QY 792 CATTAATTCAGACAGATCTTGGGCCACGAAAGCTGC---CCAGGC----- 839  
 DB 736 ---CysLeuProSerGlyAsnSerGlnGlyThrProSerCysHisProGlyArgTyrThr 754  
 QY 840 -----TTAGCTGCAGCCCG 854  
 DB 755 LeuCysLeuGlyAlaArgThrLeuGlySerSerTyrThrArgCysLeuAlaGlyThrPro 774  
 QY 855 TCATGAGAGAAAAGCTCAGAGAGAAAAGCATCTGTGGAGATGGACACTCCACCGAG 914  
 DB 775 His---GlnLysProLeuGlnHisSerArgSerCysSerGlnLeuGlyArgTyrSer 793  
 QY 915 GAGACACTGTGTGGCTCCAGCGGTTTCGATCTCAGTCCAGCCAGGACAGACAGCC 974  
 DB 794 ValAlaProAlaMetAlaPro-----ProGlnAsnSerTyrThrPhe 807  
 QY 975 ACTCCTCAGCTCTGTCCGAAACAGTACCCGACTGGCCAGCCAGGACATGTTGACCATC 1034  
 DB 808 PheProProSerLeuProLeuSerThrProArgProSerProLeuCysAsnLeuGlnAla 827  
 QY 1035 CCACCCCATGGAGCTCATCAGGAGAAAGACTAAGTCAGAGAGATCCCTCTGACCTTA 1094  
 DB 828 HisProTyrAlaLeuThrProLeuProHisLeu----- 838  
 QY 1095 CAGTTCCCTCTCTCTCCCTGCGAGCTTGATCTTGGGCCCT 1133  
 DB 839 -----LeuProGlnProLeuCysLeuGlyPro 847

RESULT 2  
 US-08-982-412-2  
 ; Sequence 2, Application US/08982412  
 ; Patent No. 5958729  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SOPPET, DANIEL R  
 ; APPLICANT: LI, YI  
 ; APPLICANT: ROSEN, CRAIG A  
 ; APPLICANT: ROSEN, STEVEN M  
 ; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
 ; STREET: 9410 KEY WEST AVENUE  
 ; CITY: ROCKVILLE,  
 ; STATE: MD  
 ; COUNTRY: US  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/982,412  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BROOKES, ANDERS A  
 ; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PF181PCT2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8439  
 ; TELEFAX: (301) 309-8439  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 884 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-982-412-2

Alignment Scores:  
 Pred. No.: 0.00518 Length: 884  
 Score: 114.00 Matches: 93

Percent Similarity:	30.02%	Conservative:	31
Best Local Similarity:	22.52%	Mismatches:	12
Query Match:	2.53%	Indels:	16
DB:	2	Gaps:	51

000 2/08-4 (1-2546) X US-08-982-412-2 (1-884

[illegible][illegible]

```

RESULT 3
5168049-3
: Patent NO. 5168049
: APPLICANT: MEADE, HARRY M.;GARWIN, J
: TITLE OF INVENTION: PRODUCTION OF S
: POLYPEPTIDES
: NUMBER OF SEQUENCES: 6
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/185,322
: FILING DATE: 21-APR-1988
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 656, 873
: FILING DATE: 02-OCT-1984
: SEQ ID NO:3
: LENGTH: 183
5168049-3

```

Alignment Scores  
Pred. No.:  
Score:  
Percent Similarity:  
Best Local Similarity:  
Query Match:  
DB:

OS-09-806-2/6A-4 (1-2546) x 5168049-3 (1-183)

```

675 ACACGACCTGGACACTGTGTTCAAGAAAGCGCTCCGG----TGCTCAAAAATGGCA 731
    ||| ||||| |||||
    25 ThrProAlaGthrArgArgProArgSerArgProProArgProAlaSerProAlaPro 44
732 -----TTCGCCGCCGACCTTGGAGATCCCAACCTCTCATGTGGCCCTAATGT 782
    ||| ||||| |||||
    45 GlyThrThrSerSerAlaArgProSerSer***ProArgAlaProThrAlaPro**** 63
783 CACCACTGACATTTAATTCACAAACAGAGATCTCGGGCCACCAAGCTGGCCGCGTTA 842
    :|| :||| :||| :|||
    64 ProGluProThrSerArgProSerAlaThrProArgAlaAlaThrSer***ProValAl 83
843 GCTGGCAGCGCCGTATGAGAGAAAAGCTCAGAGAAAGCAGTCTGTTGAGAGATGCA 902
    :||| :||| :|||
    84 ThrThrAlaProAlaGProProThrAlaAlaAlaAlaProProSerValGlyArgTrpProGly 103
903 CAGTCCACCCAGGGAGACA-----CCGCGGGGCGTCCAGGGTTTCGATCTC 950
    :|| :|| :||| :|||
    104 ArgIleThrThrAlaThrAlaThrProThrProArgProAlaGlyAla---AlaSerThrSerAla 122
951 ACCTCAGCAGGAGGAGAGACA-----GCCACTCCTCAGGCGCTGCGAACAAGTACC 1003
    :||| :||| :|||
    123 AlaProAlaGArgGlySerThrProSerArgLys***ProProAlaProProArgProThr 142
1002 CCGACTGGCCACCGGAGACATGTTGACCAATCCA-----CCCAT 1044
    ||| ||| ||||| |||
    143 Pro---GlySerProArgTrpSerAlaThrThrProSerProArg***SerArgProPro 161

```

```

QY 1044 GCGAGCTCATCAAGGAGAAAGTAACTAGTCAGAGAGTCCCTCTGACCTTACAGTTCC 1103
DB 162 PROPROSERThrArg-----ArgArgArgProAlaSerThrThrAlaThrArg 177
QY 1104 TCCTCTCCCTGCAGC 1118
DB 178 SerThrProPheSer 182

RESULT 4
US-08-841-483-6
Sequence 6, Application US/08841483B
Patent No. 5976875
GENERAL INFORMATION:
APPLICANT: Prescott, Steven M.
APPLICANT: Bunting, Michaeline
APPLICANT: Tang, Wen
APPLICANT: Topham, Matthew
TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
FILE REFERENCE: 2037.2.1a
CURRENT FILING DATE: 1997-04-22
EARLIER APPLICATION NUMBER: 60/016,210
EARLIER FILING DATE: 1996-04-22
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ. ID NO 6
LENGTH: 1117
TYPE: PRT
ORGANISM: Homo sapiens
US-08-841-483-6

Alignment Scores:
Pred. No.: 0.0141 Length: 1117
Score: 110.50 Matches: 89
Percent Similarity: 31.818 Conservative: 36
Best Local Similarity: 22.658 Mismatches: 128
Query Match: 2.46% Indels: 141
Gaps: 19

US-09-806-276a-4 (1-2546) x US-08-841-483-6 (1-1117)
QY 513 CGCTTGAGACTTTCGCCACACCATTCAGGAGAGGCGCCAGCAGATGCTTGTG 572
DB 22 ArgProSerSerValAlaLeuProThrGlyLysAlaAlaArgArgSerPro----- 38
QY 573 GAGATATTTCTTCTTCAAGGAGACTACGAGCTTTTACCTGGAAACAGAGAAAGCAGC 632
DB 39 -----AlaGlyGlnAlaSerSerSerLeuAlaGlnAlaArgArgSer 52
QY 633 ACCTGGCCAGTTCCTGGCAGTAATGAGTTCTTCGGGCGCAAGCAGCAGCTGCG 692
DB 53 Ser---AlaGlnLeuGlnGlyCysLeuLeuSerCysGlyValAlaArgAlaGlnLysSer--- 70
QY 693 TGTTCACAGAAAGC-----CCTCCCGGCTGCTCA----- 722
DB 71 -----SerArgArgArgSerSerThrValProProSerCysAsnProArgPheIleValAsp 89
QY 723 AAAATGCCATCTCCCTC-CCGACCATTTGAGAGATCCCAAGCTCATGTTGCCCTTATG 781
DB 90 LysValLeuThrProGlnProThrThrValGlyAlaGlnLeuLeuGlyAlaProLeuLeu 109
QY 782 TCACACAGTACATTAAATTCACAAACAGAGTCTCTTCGGCCAGCAAAAGCTGCCAGGCTT 841
DB 110 LeuThrGlyLeuValGlyMetAsnGlnLys----- 119
QY 842 ACCTCGAGCCCGTCATGAGAGAAAGCTCAGAGAAAGCAGTCTGTTGAGAAATGG 901
DB 120 -----GlnGlyValGlnGlnAspValValAlaGlnAlaSerSerAlaIleGlnProGly 137
QY 902 ACAGTCACAC-----AGGAGAGACACCTGCGGGCTTCAGACGGTTCATCTCAGTCC 955
DB 138 Thr-LysThrProGlyProProProProArgGlyAlaGlnProLeuLeuProLeu---Pr 156

QY 956 AGCCAGGCGAGAGACAGCCACCTCTCCAGCCTGCC----- 992
DB 156 GArgGlyValAlaArgAlaSerSerHisCysCysProAlaAspAlaValLysHisAl 176
QY 993 -----AACAGTACCCCGAGTGGCCAG-- 1013
DB 176 AleuThrGlyLeuHisGlyTyrTyrArgArgLeuSerGlnArgArgProSerGlyGlnH 196
QY 1014 -----CCGAGGACATGTTGACCATCC 1036
DB 196 SProGlyProGlyLysArgArgAlaSerGlyThrThrAlaGlyThrMetLeuProThrAr 216
QY 1037 ACCCCATGCGAGCTCATCAAGGAGAAAGTAACTAGTCAGAGAGTCCCTCTGACCTTACA 1096
DB 216 GValArgProLeuSerArgArgArgGlnValAlaLeuAlaGlyGlyAlaAlaGlyProGl 236
QY 1097 G----- 1097
DB 236 AlaThrSerAlaLeuLeuAlaLysAlaIleThrLysSerGlyLeuGlnHisLeuAlaPr 256
QY 1098 -----GTTCCCTCTCTCCCTGC----- 1115
DB 256 OProProThrProGlyAlaProCysSerGluSerGluArgGlnIleArgSerThrVa 276
QY 1116 -----AGCTTGATCTTGGGCGCTCATTTTGGAGTGAAGGCTGATGTAATG 1162
DB 276 LAspThrSerLeuSerAlaThrThrGlyGlnHisIleThrPheGlu----- 291
QY 1163 GATAAAATAATAGTAAAGATGCCACTTTTTCCTTTGGGTAAGATG-----AA 1216
DB 292 -----ThrAsnValSerGlyAspPheCysTyrValGlyGlnGlnIleTyrCysValAl 308
QY 1217 AACAACTAACACAGTTCGAAGAGAGGCGCTTCGAGCTGATTTGACGTTTGTGTT 1276
DB 308 ArgMetLeuLysSerValSerArgArgLysCysAlaAlaCys----- 322
QY 1277 GGGTTTCTAAATAATATTTCTTACAAAGTATTTTTCCTGTATGCGCTTTGCAA 1336
DB 323 -----LysIleValValHis-----ThrProCysII 331
QY 1337 AACAACTTAAAGAAAA-----AACACAAAGCAAAACCTATCTTGGCAAAAGGAGCTGA 1393
DB 331 eGlnGlnLeuGlnLysIleAsnPheArgCysLysProSerPheArgGlnLysSerAr 351
QY 1394 GTCA-----GAGCCCATTTTC--AGGAGCATTTGG 1421
DB 351 GAsnValArgGluProThrPheValAlaArgHisHisTrp 363

RESULT 5
US-09-382-911-6
Sequence 6, Application US/09382911
Patent No. 6221658
GENERAL INFORMATION:
APPLICANT: Prescott, Steven M.
APPLICANT: Bunting, Michaeline
APPLICANT: Tang, Wen
APPLICANT: Topham, Matthew
TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
FILE REFERENCE: 2037.2.1a
CURRENT FILING DATE: 1999-08-25
CURRENT FILING DATE: 1999-08-25
CURRENT FILING DATE: 1997-04-22
PRIOR APPLICATION NUMBER: 08/841,483
PRIOR FILING DATE: 1997-04-22
PRIOR APPLICATION NUMBER: 60/016,210
PRIOR FILING DATE: 1996-04-22
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ. ID NO 6
LENGTH: 1117
TYPE: PRT
ORGANISM: Homo sapiens

```

US-09-382-911-6

## Alignment Scores:

Pred. No.: 0.0141 Length: 1117  
 Score: 110.50 Matches: 89  
 Percent Similarity: 31.81% Conservative: 36  
 Best Local Similarity: 22.65% Mismatches: 128  
 Query Match: 2.46% Indels: 141  
 DB: 4 Gaps: 19

US-09-806-276a-4 (1-2546) x US-09-382-911-6 (1-1117)

```

QY 513 CGCTGGAGACTTTCGCCACACCATTCAGTGGCAAGAGGCGCAGCATGCTTTG 572
Db      |||      |||      |||      |||      |||      |||      |||
QY 22 ArgProSerSerValGlyLeuProThnGlyAlaArgArgArgSerPro----- 38
QY 573 GAGATATTCTCTTCACAGACTACGACTTTTACTGTGAAACAGAGAAAGCAC 632
Db      |||      |||      |||      |||      |||      |||      |||
QY 39 -----AlaGlyGlnAlaSerSerSerLeuAlaGlnArgArgSer 52
QY 633 ACCTGGGCCAGTTCCTGGGACATATGAGTTCTCCGGCCACACGACCTCGACTGTG 692
Db      |||      |||      |||      |||      |||      |||      |||
QY 53 Ser--AlaGlnLeuGlnGlyLeuProThnGlyAlaArgArgArgSerPro----- 70
QY 693 TGTTACAGAAAGC-----CCTCCCGGCTCTCA-----722
Db      |||      |||      |||      |||      |||      |||      |||
QY 71 --SerArgArgArgSerSerThnValProProSerCysInProArgPheLeuValAsp 89
QY 723 AAATGGCATCTCCCTC-CGACCATTTGGAGATCCGACGCTCATGTTGCCATTATG 781
Db      |||      |||      |||      |||      |||      |||      |||
QY 90 LysValLeuThnProThnProThnValGlyAlaGlnLeuLeuValAlaProLeu 109
QY 782 TCACCAAGTACATTAATTCGAAACAGAGAGTCTTCGGGCCAGCAAGCTGCCAGCTT 841
Db      |||      |||      |||      |||      |||      |||      |||
QY 110 LeuThnGlyLeuValAlaGlyMetAsnGlnGlu-----119
QY 842 AGCTGGGAGCCCTCATTCGAGGAGAAAGCTCAGAGAGAAAGCAAGTCTTTGGAGATGG 901
Db      |||      |||      |||      |||      |||      |||      |||
QY 120 -----GlnGlyValAlaGlnGlnSerValAlaAlaGlnAlaSerSerAlaGlnProGly 137
QY 902 ACAGTCCACC-----AGGAGACACCTCGTGGGCTTCGACGCTCATGTCAGTCC 955
Db      |||      |||      |||      |||      |||      |||      |||
QY 138 Thr-LysThnProGlyProProProProProGlnGlyAlaGlnProLeuProLeu--Pr 156
QY 956 AGCCAGGAGAGAGACAGCCATCTCCAGCCTGTCCG-----992
QY 156 oArgTyValAlaArgArgAlaSerSerHisCysProAlaAspAlaValTyrAspHisAl 176
QY 993 -----AACAGTACCCCGACtGGCCAG--1013
QY 176 AlaLeuThnGlyLeuHisGlyTyrTyrArgArgLeuSerGlnArgArgProSerGlyGlnH 196
QY 1014 -----CCGAGAGACATGTTTGACCATGCC 1036
QY 196 sProGlyProGlyGlyAlaArgArgAlaSerGlyThnThnAlaGlyThnMetLeuProThnAr 216
QY 1037 ACCCCATGGGAGCTCATCAAGGAAAGACTAAGTACAGAGAGTCCCTCTGACCTTACA 1096
QY 216 gValAlaArgProLeuSerArgArgArgGlnValAlaLeuArgArgLysAlaAlaGlyProG 236
QY 1097 G-----1097
QY 236 nAlaTrpSerAlaLeuLeuAlaLysAlaAlaLeuThnLysSerGlyLeuGlnHisLeuAlaPr 256
QY 1098 -----GTTCCCTCTCTCCCTGC-----1115
QY 256 oProProProThnProGlyAlaProCysSerGlnSerGlnArgGlnIleArgSerThnVa 276
QY 1116 -----ACCTTGATCTTGCGCCCTCAGCTTTTGATGAGAGCTGTAATGTAATG 1162
QY 276 lAspTrpSerGlnSerAlaAlaThnTyrGlyGlnHisIleTrpPheGlu-----291
QY 1163 GATAAATAATAGTACAGAGATGCCAAACTTTTCTCTTGCGGTAAAGGTAC-----AA 1216
  
```

```

Db 292 -----ThnAsnValSerGlyAspPheCysTyrValGlyGlnGlnTyrCysValAl 308
QY 1217 AAACAACACTAACACAGTGTGAAGAGAGCGCTCCGAGCTGTATTTGACGTTTGTTG 1276
Db      |||      |||      |||      |||      |||      |||      |||
QY 308 aArgMetLeuLysSerValSerArgArgLysCysAlaAlaCys-----322
QY 1277 GGGTTTTCAAATAATATCTTACAAAGATATTTTACTGCTGTATGCGCTGTTTGCA 1336
Db      |||      |||      |||      |||      |||      |||      |||
QY 323 -----LysIleValValHis-----ThnProCysI 331
QY 1337 AAACAATTAGAAAAA--AACACAACAAGCAAAACCTATCTTGGCAAAAAAGGAGTGA 1393
Db      |||      |||      |||      |||      |||      |||      |||
QY 331 eGlnGlnLeuGlnGlyLysAsnPheArgCysLysProSerPheArgGlnSerGlySerAr 351
QY 1394 GTCA-----GAGCCCATTTTC--AGAGAGCATTTG 1421
Db      |||      |||      |||      |||      |||      |||      |||
QY 351 gAsnValAlaArgGlnProThnPheValAlaGlnHisIleTrp 363
  
```

RESULT 6  
 PCT-US93-03027-1  
 ; Sequence 1, Application PC/TUS9303027  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LEONARD, WARREN; TOLEDANO,  
 ; APPLICANT: MICHEL  
 ; TITLE OF INVENTION: CONTROL AND/OR  
 ; TITLE OF INVENTION: PREVENTION OF BINDING OF NF- B/REL/DORSAL  
 ; TITLE OF INVENTION:  
 ; NUMBER OF SEQUENCES: 9  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MORGAN & FINNEGAN  
 ; STREET: 345 PARK AVENUE  
 ; CITY: NEW YORK  
 ; STATE: NEW YORK  
 ; COUNTRY: USA  
 ; ZIP: 10154  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: FLOPPY DISK  
 ; COMPUTER: IBM PC COMPATIBLE  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WORDPERFECT 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/03027  
 ; FILING DATE: 19930401  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/862,987  
 ; FILING DATE: 06-APR-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: DOROTHY R. AUTH  
 ; REGISTRATION NUMBER: P-36,434  
 ; REFERENCE/DOCKET NUMBER: 2026-4010 PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-758-4800  
 ; TELEFAX: 212-751-6849  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 550  
 ; TYPE: AMINO ACID  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: human  
 ; STRAIN:  
 ; INDIVIDUAL ISOLATE:  
 ; DEVELOPMENTAL STAGE:  
 ; HAPLOTYPE:  
 ; TISSUE TYPE:  
 ; CELL TYPE:  
 ; CELL LINE:  
 ; ORGANELLER:  
 ; FEATURE:



```

SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapien
TISSUE TYPE: Liver CDNA sequence
US-07-864-004B-4

Alignment Scores:
Pred. No.: 0.0252 Length: 2332
Score: 110.00 Matches: 79
Percent Similarity: 38.51% Conservative: 55
Best Local Similarity: 22.70% Mismatches: 122
Query Match: 2.44% Indels: 92
DB: Gaps: 17

US-09-806-276a-4 (1-2546) x US-07-864-004B-4 (1-2332)

QY 285 CCACCTCTGAGCCACTTGAAGAAAATGATGTGACAGTTCATCAAAAAGATTGACGA 344
DB 749 ProSerThrArgLysGlnPheAsnAlaThrThrIlePro-GluAsnAspIleGluLys 768
QY 345 A-----CATATACCATCTGAGAAAGTGGCCCTTTC 377
DB 768 sThrAspProTyrPheAlaHisArgThrProMetProLysIleGlnAsnValSerSerSe 788
QY 378 TCCCGCTTCAAAATAGACATTCTCAAAATTCGCAAGAGCCCAATTTACG 437
DB 788 r-----AspLeuLeuMetLeuLeuArgLysInsr-ProThrProHisGlyL 803
QY 438 TGAAGCAGCCATATACAGAAAGAAAGAAATTTAACTGAGGACATTCGTCTCGT 497
DB 803 eSerLeuSerAspLeuGlnGluAlaLysTyrGluThrPheSerAspAspProSerProG 823
QY 498 ATATGATCAGT-----CCCCCGCTTGAGAGACTTTCGCCACACCATCCACA 542
DB 823 lYAlAlIleAspSerAsnAsnSerLeuSerGluMetThrHisPheArgProGlnLeuHis- 842
QY 543 TTGGCAAGAGGCCAGCAGATGCTTGGAGATATTTCCTT----- 586
DB 843 -----HisSer-----GlyAspMetValPheThrProGlnSerGlyL 855
QY 587 -----CTTCAAGGAAGTACGAGCTTTTACCTGGGAAACGAGAGAAACACACCTGGGCC 641
DB 855 eugLInLeuArgLeuAsnGlnLysLeu-----GlyThrThrAlaAlaThrGlnLeuLysL 873
QY 642 ACTTCCTTGCGCANTATGATCTTCCGCGGCCAACAGCAGCTCGGACTGTGTTCAAG 701
DB 873 yLLeuAsp-----PheLysValSerSerThrSerAsnAsnLeuIleSerT 888
QY 702 AAACCCCTCCCGGTGCTC-----AAAAATGCCATCTCCCTCCGACCATTTG 749
DB 888 hTlIeProSerAspAsnLeuAlaAlaGlyThrAspAsnThrSerSerLeuGly----- 905
QY 750 GAGGATCCCAAGCTCTCATGTTGACCTTATTGTCACTGACATTTAATCCAAACAG 809
DB 906 -----ProProSerMetProValHisTyrTrsPserSerLeuLeu 918
QY 810 AGTCC-----TTGCGGCCAGCAAAAGCTGCCAGGCTT--AGCTGGAGCCCGTCATGG 860
DB 918 sPThrThrLeuPheGlyLysLysSerSerProLeuThrGlnSerGlyGlyProLeuSerL 938
QY 861 AGGAAAAGCTCAGAGAGAAAGCACTGTGTGGAGATGGAGACGTCCACGAGGAGACA 920
DB 938 eusergLInGluAsnAsnAspSerLysLeuGlnSerGlyLeuMetAsnSerGlnGlnLus 958
QY 921 CCTGCTGGGCTCCAGCGGTTCTGCATCTCAGTCAGC-----CAAGCGAGACACA 971

```

```

DB 958 eSerTyrGlyLysAsnValSerThrThrGlnSerGlyArgLeuPheLysGlyLysArgA 978
QY 972 GCCACTCTCCAGCCCTGTCCGAAACAGTACCCCGACTGCGCCAGAGACATGTTGACC 1031
DB 978 lAhISgLyProAlaLeuLeu-----ThLysAspAsnAlaLeuP 991
QY 1032 ATCCACCCCATGCGAGCTCATCAAGGAAAGACTAAGTCAAGAGAGTCCCTCTGAC- 1090
DB 991 heLysValSerLleSerLeuLeuLysThrAsnLysThrSerAsnAsnSerAlaThrAsnA 1011
QY 1091 -----CTTACAGTTCCTCCCTCTCCCTGAGACTTGATCTGGGCCCTCATTT 1139
DB 1011 rGlyThrHisLaspolyPro-----SerLeuLeuIleGlnAsnSerProSerValT 1029
QY 1140 TGGATGAGTCTCGAT----- 1156
DB 1029 rPGLInAsnIleLeuGlnSerAspThrGlnPheLysValThrProLeuIleHisAspA 1049
QY 1157 -----GTATGATATAAAT 1171
DB 1049 rGmetLeuMetAspLysAsn 1055

RESULT 8
US-08-251-937A-4
; Sequence 4, Application US/08251937A
; Patent No. 5583209
;
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; APPLICANT: Runge, Marshall S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/251,937A
; FILING DATE: 31-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pratt, John S.
; REGISTRATION NUMBER: 29,476
; REFERENCE/DOCKET NUMBER: EMU106DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6367
; TELEFAX: 404-815-6555
; INFORMATION FOR SFO ID NO: 4:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver CDNA sequence
;
; US-08-251-937A-4

```

Alignment Scores:  
 Pred. No.: 0.0252 Length: 2332  
 Score: 110.00 Matches: 79  
 Percent Similarity: 38.51% Conservative: 55  
 Best Local Similarity: 22.70% Mismatches: 122  
 Query Match: 2.44% Indels: 92  
 DB: 1 Gaps: 17

US-09-806-276a-4 (1-2546) x US-08-251-937a-4 (1-2332)

```

OY 285 CCACCTCTGAGCCACTTGAGAAATGATGTGACAGCTCTATCAACAAAGATTGACAA 344
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 749 ProSerThrArgInLysGlnPheAsnAlaThrThrIlePro-GluAsnAspIleGluJ 768
    345 A-----CATATACCATGTGTGAGAAAGTGGCCCTTTC 377
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 768 sThrAspProTrpPheAlaHisArgThrProMetProLysIleGlnAsnValSerSe 788
    378 TCCCGGCTGCAAAATAGACATTCTCAAAATTCGACGACGACGACGACGACGACGAC 437
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 788 r-----AspleuMetLeuLeuArgGlnSer-ProThrProHisGlyL 803
    438 TGAAGACGCAATTAACGAAAGAAATTTAACTGAGGACATTTCTGTCTCTG 497
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 803 euserLeuSerAspLeuGlnGlnAlaLysTrpGluThrPheSerAspAspProSerPro 823
    498 ATATGATCACT-----CCCCGCTTGAGACATTTGCCGACACCATTCACA 542
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 823 LysAlaLeaSerAsnAsnSerLeuSerGlnMetThrHisPheArgProGlnLeuHis 842
    543 TTGGCAAGAGGGGCGACGATGTCTTGAGATATTTTCCTT----- 586
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 843 -----HisSer-----GlyAspMetValPheThrProGlnSerGlyL 855
    587 -----CTTCAAGGAGAACTACGACCTTTTACCTGGAACGAGGAAAGACACACTGGGCC 641
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 855 eugInLeuArgLeuAsnGlnLysLeu-----GlyThrThrAlaAlaThrGlnLeuLysL 873
    642 AGTTCCCTGGGCAATAGATTTCTCCGGGCGCAACGACCTGGGACCTGTGTTCACAG 701
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 873 ystLeuAsp-----PheLysValSerSerThrSerAsnAsnLeuIleSerL 888
    702 AAAGCCCTCCCGGCTGCTC-----AAAATGCCATCTCCCTCCGACCATTTG 749
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 888 hrIleProSerAspAsnLeuAlaLysGlyThrAspAsnThrSerSerLeuGly----- 905
    750 GAGGATCCCAAGCTCTCATGTTGCCCTTATTGTCACGATGACATTTAATTCACAAACAG 809
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 906 -----ProProSerMetProValHisTyrAspSerGlnLeuA 918
    810 AGTCC-----TTGGGGCGACGAAAGCTGCCAGGCTT---AGCTGGAGCCCGCTCATG 860
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 918 sPrThrThrLeuPheGlyLysLysSerSerProLeuThrGlnSerGlyGlyProLeuSerL 938
    861 AGGAAAGAGCTCAGAGAAAGAGCTGTGTGGAGATGGAGACAGTCCACAGGAGAC 920
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 938 euserGlnGlnAsnAsnAspSerLysLeuLeuGlnSerGlyLeuMetAsnSerGlnGln 958
    921 CTTGTGGGGCTCCAGCGCTTCTGCATCTCAGTCCAC-----CAAGGACAGAC 971
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 958 erSerTrpGlyLysAsnValSerSerThrGlnSerGlyArgLeuPheLysGlyLysArg 978
    972 GCGACCTCCGACGCTGCGCAACAGTACCCCGACTGGCCAGCGAGACATTTTTCAGC 1031
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 978 LHisGlyProAlaLeuLeu-----ThrLysAspAsnAlaLeuP 991
    1032 ATCCACCCCATGCGAGCTCAAGGAGAAAGACTAAGTACGAGAGACTCCCTCTCTGAC- 1090
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 991 heLysValSerLleSerLeuLeuLysThrAsnLysThrSerAsnAsnSerAlaThrAsnA 1011
    1091 -----CTTACAGGTTCCCTCTCTCCCTCGACGCTTGAGTCTGGGCGCTTCATTT 1139
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1011 rgLysThrHisIleAspGlyPro-----SerLeuLeuIleGlnAsnSerProSerValT 1029
  
```

```

OY 1140 TGGATGAGTGCTCTCAAT----- 1156
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1029 rpgInAsnIleLeuGlnSerAspThrGlnPheLysLysValThrProLeuIleHisAspA 1049
    OY 1157 -----GTAATGATGATAAAT 1171
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 1049 rgMetLeuMetAspLysAsn 1055
  
```

RESULT 9

US-08-212-133A-2

; Sequence 2, Application US/08212133A

; Patent No. 5663060

; GENERAL INFORMATION:

; APPLICANT: Koller, John S.

; APPLICANT: Runge, Marshall S.

; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESS: Kilpatrick & Cody

; STREET: 100 Peachtree Street

; CITY: Atlanta

; STATE: Georgia

; COUNTRY: US

; ZIP: 30303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/212,133A

; FILING DATE: March 11, 1994

; CLASSIFICATION: A35

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 07/864,004

; FILING DATE: 07-APR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Pabst, Patricia L.

; REGISTRATION NUMBER: 31,284

; TELEPHONE: 404-572-6555

; TELEFAX: 404-572-6555

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2332 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: YES

; ANTI-SENSE: NO

; FRAGMENT TYPE: N-terminal

; ORIGINAL SOURCE:

; ORGANISM: Homo sapien

; TISSUE TYPE: Liver cDNA sequence

US-08-212-133A-2

Alignment Scores:  
 Pred. No.: 0.0252 Length: 2332  
 Score: 110.00 Matches: 79  
 Percent Similarity: 38.51% Conservative: 55  
 Best Local Similarity: 22.70% Mismatches: 122  
 Query Match: 2.44% Indels: 92  
 DB: 1 Gaps: 17

US-09-806-276a-4 (1-2546) x US-08-212-133A-2 (1-2332)

```

OY 285 CCACCTCTGAGCCACTTGAGAAATGATGTGACAGCTCTATCAACAAAGATTGACAA 344
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 749 ProSerThrArgInLysGlnPheAsnAlaThrThrIlePro-GluAsnAspIleGluJ 768
  
```



345 A-----CATATACATCTGTGAGAAAGTGGCCCTTTC 377  
 Db 768 sThraspProTprhealAhlsArgThrProMetProLysIleGlnAsnValSerSeSe 788  
 QY 378 TCCCGCTTGCAAAATAGCATTTCCAAATTCGCAAGGAGCAAGCCCAATTATAC 437  
 Db 788 r-----AspleuMetIleuArgGlnSer-ProThrProhlsGlyL 803  
 QY 438 TGAAGAGCCCAATACAGAAAGAAAGAAATTAAGTGAAGGACATTCCTGCTG 497  
 Db 803 euserLeuSerAspleuGlnGlnLysIleThrPheSerAspAspProSerProG 823  
 QY 498 ATATGATCAGT-----CCCCCGCTTGAGAGCTTGGCCGACACCAATCCACA 542  
 Db 823 lYalAlleAspSerAsnAsnSerLeuSerGluMetThrHisPheArgProGlnLeuHis- 842  
 QY 543 TTGGCAAGAGGCCGACGATGCTTGTGGAGATATTCCTTT-----HisSer 586  
 Db 843 -----HisSer-----GlyAspMetValPheThrProGluSerGlyL 855  
 QY 587 -----CTTCAAGGAACTACGAGCTTTTACTGTGAAACAGAGAAAGCAACCTGGGCC 641  
 Db 855 euGlnLeuArgLeuAsnGlnLysLeu-----GlyThrThrAlaAlaThrGlnLeuLysL 873  
 QY 642 AGTCCCTGGCATATAGATTCTTCCGGGCAAGCAAGCCCTGGGACTGTGTTCACAG 701  
 Db 873 yLeuAsp-----PheLysValSerSerThrSerAsnLeuIleSer 888  
 QY 702 AAAGCCCTCCCGGCTCTC-----AAATGCCATTCCTCCCGGACCAATG 749  
 Db 888 hrllePserAspAsnLeuAlaGlyThrAspAsnThrSerSerLeuLys----- 905  
 QY 750 GAGATGCCAAGCTCATCTGTCCTTATGTCACAGTACGATTAATTCACAAACAG 809  
 Db 906 -----ProPserMetProValHisIleTyrAspSerGlnLeuA 918  
 QY 810 AGTCC-----TTGGGCGCAGCAAAAGCTGCCAGGCTT--AGCTGGCAAGCCGTCATG 860  
 Db 918 sphThrLeuPheGlyLysLysSerSerProLeuThrGlnSerGlyGlyProLeuSerL 938  
 QY 861 AGGAAAGCTCAGAGAAAGCAATCTGTGGAGATGGACAGATCCACAGGAGACA 920  
 Db 938 euserGlnLysAsnAsnAspSerLysLeuGlnSerGlyLeuMetAsnSerGlnLys 958  
 QY 921 CCGCGGGGCTCCAGCGGTTCTGATCTCATGTCAGTCCAGC-----CAAGGACAGACA 971  
 Db 958 eSerTrpLeuLysAsnValSerSerThrGlnSerIleThrGlnPheLysGlyLysArg 978  
 QY 972 GCCACTCTCCAGCTGCGGACAGTACCCGACACTGGCCAGGACGATGTTGACC 1031  
 Db 978 lAhlsGlyProAlaLeu-----ThrLysAspAsnAlaLeuP 991  
 QY 1032 ATCCACCCCATGCGAGCTCATCAAGGAAAGCTAAGTACAGAGAGTCCCTCTGAC- 1090  
 Db 991 helysValSerIleSerLeuLysThrAsnLysThrSerAsnAsnSerAlaThrAsnA 1011  
 QY 1091 -----CTTAAGAGTCCCTCCCTGCTCCGCTGACACTTGGAGCTTGGCCCTCACTT 1139  
 Db 1011 rglYstThrHisIleAspGlyPro-----SerLeuLeuIleGlnAsnSerProSerVal 1029  
 QY 1140 TCGATGAGGCGTCAAT----- 1156  
 Db 1029 rplAsnIleLeuLysSerAspThrGlnPheLysLysValThrProLeuIleHisAspA 1049  
 QY 1157 -----GTAATGATAAAT 1171  
 Db 1049 rGmetLeuMetAspLysAsn 1055

APPLICANT: Emory University  
 TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
 NUMBER OF SEQUENCES: 12  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Kilpatrick & Cody  
 STREET: 1100 Peachtree Street, Suite 2800  
 CITY: Atlanta  
 STATE: Georgia  
 COUNTRY: US  
 ZIP: 30309  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/474,503  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Pratl, John S.  
 REGISTRATION NUMBER: 29,476  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 404-815-6550  
 TELEFAX: 404-815-6555  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2332 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHEtical: YES  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: N-terminal  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapien  
 Tissue Type: Liver CDNA sequence  
 US-08-474-503-2  
 Alignment Scores:  
 Pred. No.: 0.0252  
 Score: 110.00  
 Percent Similarity: 38.51%  
 Best Local Similarity: 22.70%  
 Query Match: 2.44%  
 DB: 1  
 Gaps: 17  
 US-09-806-276a-4 (1-2546) x US-08-474-503-2 (1-2332)  
 QY 285 CCACTTCTGAGGCACTTGAAGAAATGATGTGACAGTCTTCAAAAAAGATTGAGA 344  
 Db 749 ProSerThrArgGlnGlnPheAsnAlaThrThrIlePro-GluAsnAspIleGlyL 768  
 QY 345 A-----CATATACATCTGTGAGAAAGTGGCCCTTTC 377  
 Db 768 sThraspProTprhealAhlsArgThrProMetProLysIleGlnAsnValSerSeSe 788  
 QY 378 TCCCGCTTGCAAAATAGCATTTCCAAATTCGCAAGGAGCAAGCCCAATTATAC 437  
 Db 788 r-----AspleuMetIleuArgGlnSer-ProThrProhlsGlyL 803  
 QY 438 TGAAGAGCCCAATACAGAAAGAAAGAAATTAAGTGAAGGACATTCCTGCTG 497  
 Db 803 euserLeuSerAspleuGlnGlnLysIleThrPheSerAspAspProSerProG 823  
 QY 498 ATATGATCAGT-----CCCCCGCTTGAGAGCTTGGCCGACACCAATCCACA 542  
 Db 823 lYalAlleAspSerAsnAsnSerLeuSerGluMetThrHisPheArgProGlnLeuHis- 842  
 QY 543 TTGGCAAGAGGCCGACGATGCTTGTGGAGATATTCCTTT-----HisSer 586

Fri Jul 25 09:59:19 2003

us-09-806-276a-4.ra1

Page 10

Db 843 -----HisSer-----GlyAspMetValPheThrProGluSerGlyL 855  
OY 587 -----CTTCAAGGAGGACTACGAGCTTTTACCTGGAGAAACAGAGAAAGACACCTGGGCC 641  
Db 855 euglnleuargleuasnlglylsleu-----GlyThrThralalaThrGluLeuYsL 873  
OY 642 AGTTCCTGGGCAATATGAGTCTTCGGGCAACAGACCTCGGACCTGCTGTACAG 701  
Db 873 yslauAsp-----PheYsValSerSerThrSerAsnSerleuIleSerT 888  
OY 702 AAACGCTCCCTCCGCTGCTC-----AAATGCAATCTCCCTCCGACCATATG 905  
Db 888 hrlleProSerAspAsnleuAlaaglyThrAspAsnThrSerleuGly----- 905  
OY 750 GAGATCCCAAGCTCTCATGTTGCCCTATTGTCACAGATGATTAATTCACAGAG 809  
Db 906 -----ProSerMetProValHisThrAspSerGlnLeuA 918  
OY 810 AGTCC-----TTGGGGCCAGCAAGCTGCCAGGCTT---AGCTGGAGCCCGTCATG 860  
Db 918 sprThrleuPheGlyLysSerSerProleuThrGluSerGlyLProleuSerL 938  
OY 861 AGGAAAGCTCAGAGAAAGACAGTCTGTGGAGAAATGAGGACAGCTCCAGGAGACA 920  
Db 938 euserGluGluAsnAspSerLysleuLeuGluSerleuMetAsnSerGlnLus 958  
OY 921 CCTGTGGGGCTCCAGGGGTTTCGATCTCATCTCAGC-----CAAGCCAGACAGA 971  
Db 958 erSerTrpGlyLysAsnValSerSerThrGluSerGlyArgleuPheYsGlyLysArg 978  
OY 972 GGCACCTCTCCAGCCTCTCCGACAGTACCCGACGCGCAGCCAGGACATGTTGAC 1031  
Db 978 lahlsGlyProAlaLeuLeu-----ThrLysAspAsnAlaLeuP 991  
OY 1032 ATCCCAAGGCTCAGGCTCATCAAGGAAAGACATAAGTACAGAGGCTCTCTGAC- 1090  
Db 991 helysValSerleuSerleuLeuYsThrAsnLysThrSerAsnAsnSerAlaThrAsnA 1011  
OY 1091 -----CTTACAGTTCCCTCCCTCCGCTGAGCTTGTGGGCTCTCATCTT 1139  
Db 1011 rglYsThrHisIleAspGlyPro-----SerleuLeuIleGluAsnSerProSerValT 1029  
OY 1140 TGGATGAGGTGCTGAAT----- 1156  
Db 1029 rpglnAsnIleleuGluSerAspThrGluPheLysValThrProleuIleHisAspA 1049  
OY 1157 -----GTATGGATAAAAT 1171  
Db 1049 rgmleuMetAspLysAsn 1055

RESULT 11  
US-08-670-707A-2  
Sequence 2, Application US/08670707A  
Patent No. 5859204  
GENERAL INFORMATION:  
APPLICANT: Lohat, John S.  
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/670.707A  
FILING DATE: 26-JUN-1996

CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: NO PCT/US94/13200  
FILING DATE: 15-NOV-1994  
PRIOR APPLICATION DATA: US 08/212,133  
APPLICATION NUMBER: 11-MAR-1994  
FILING DATE: 07-APR-1992  
PRIOR APPLICATION DATA: US 07/864,004  
APPLICATION NUMBER: 07-APR-1992  
FILING DATE: 07-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Greenlee, Lorance L.  
REGISTRATION NUMBER: 27,894  
REFERENCE/DOCKET NUMBER: 75-95F  
TELEPHONE: 303/499-8089  
TELEFAX: 303/499-8080  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2332 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHEICAL: YES  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Liver

US-08-670-707A-2

Alignment Scores:  
Pred. No.: 0.0252 Length: 2332  
Score: 110.00 Matches: 79  
Percent Similarity: 38.51% Conservative: 55  
Best Local Similarity: 22.70% Mismatches: 122  
Query Match: 2.44% Indels: 92  
Gaps: 17

US-09-806-276A-4 (1-2546) x US-08-670-707A-2 (1-2332)

OY 285 CCACCTTTCAGAGCCACTTTCAGAAATGATGACAGTTCATGCAAAAGGATTCAGAA 344  
Db 749 ProSerThrArgGlnLysGlnPheAsnAlaThrThrlePro-GluAsnAspIleGlyL 768  
OY 345 A-----CATATACATCTGTGAAGAAAGTGCCCTTTC 377  
Db 768 sthAspProThrPheAlaHisArgThrProMetProLysIleGlnAspValSerSer 788  
OY 378 TCCCGCTTGCAGAAATGACATTCGAAATTCGAAATGCCAGCCAGCCCAATTACC 437  
Db 788 r-----AspLeuLeuMetLeuLeuArgGlnSer-ProThrProHisGlyL 803  
OY 438 TGAAGCAGCCATATACAGAAAGAAAGAAATTAACAGGACATTCCTGCTCCG 497  
Db 803 euserLeuSerAspLeuGlnGlnAlaLysTyrGluThrPheSerAspAspProSerProG 823  
OY 498 AATATGATCAGT-----CCCCGCTTGGAGACTTTCGCCACACATCCACA 542  
Db 823 lYAlaIleAspSerAsnAsnSerleuSerGluMetThrHisPheArgProGlnLeuHis 842  
OY 543 TTGCAAGAGAGGCGCAGCAGATGCTTTCGAAACAGAGAGAGAGAGAGAGAGAGAG 586  
Db 843 -----HisSer-----GlyAspMetValPheThrProGluSerGlyL 855  
OY 587 -----CTTCAAGGAGGACTACGAGCTTTTACCTGGAGAAACAGAGAAAGACACCTGGGCC 641  
Db 855 euglnleuargleuasnlglylsleu-----GlyThrThralalaThrGluLeuYsL 873  
OY 642 AGTTCCTGGGCAATATGAGTCTTCGGGCAACAGACCTCGGACCTGCTGTACAG 701  
Db 873 yslauAsp-----PheYsValSerSerThrSerAsnSerleuIleSerT 888

```

QY 702 AAGCGCCCTCCCGGTC-----AAAATGCATCTCCCTCCCGACCATG 749
      ||||| ||| |||||
Db 888 hrlleProseraspnleuAlaIaIaglyThAspAnTherSerleugly----- 905
QY 750 GAGGATCCCAAGCTCATGTTGCCCTTATGTCACAGTATTAATTCAAACAG 809
      ||| ||||| |||||
Db 906 -----ProProSerMetProValHisTyrAspSerGlnLeu 918
QY 810 AGTCC-----TTGGGGCCAGCAAGACTTCCAGGCTT---AGCTCCGAGCCCGCTCATG 860
      ||||| ||||| |||||
Db 918 sphrThrLeuphegLyLysSerSerProLeuThrgLusSerGlyGlyProLeuSerL 938
QY 861 AGGAAAGAGCTGAGGAAAGACAGTCTGTGGAGATGAGACAGTCCAGGAGACA 920
      ||| ||||| |||||
Db 938 euSerGlnLusAnAspSerLyLLeuLusLusSerLyLLeuMetAnSerGlnLus 958
QY 921 CCTGTGGGGCTCCAGCGGTTCTGATCTCATGTCAGC-----CAAGCAGACACA 971
      ||||| ||||| |||||
Db 958 eSerTrpGlyLysAnValSerSerThrgLusSerGlyLysLeuPheLysGlyLysArg 978
QY 972 GCCACCTCTCCAGCTGTCGGAACAGTACCCGACTGGCCAGCCGAGACATGTTGACC 1031
      ||||| ||||| |||||
Db 978 lahslgLyProAlaLeuLeu-----ThrLysAspAnAlaLeu 991
QY 1032 ATCCACCCCATGCGAGCTCATCAGGAAAGACTAGTACAGAGAGTCCCTCTGAC- 1090
      ||||| ||||| |||||
Db 991 heLysValSerLeuLeuLysThrsAnLysThrsAnAnSerAlaThAsnA 1011
QY 1091 -----CTTACAGGTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1139
      ||||| ||||| |||||
Db 1011 rglYsthrHisLLeaspGlyPro-----SerLeuLeuLLeGlnAnSerProSerVal 1029
QY 1140 TGGATGAGGTCGCGAAT----- 1156
Db 1029 rplGlnAnLLeuGlnSerAspThrgLusLysValThrProLeuLLeHisAspA 1049
QY 1157 -----GTAATGATATAAAT 1171
      ||||| ||||| |||||
Db 1049 rgmleuMetAspLysAsn 1055

RESULT 12
US-09-037-601-2
; Sequence 2, Application US/09037601
; Patent No. 6180371
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,601
; FILING DATE: 26-JUN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004

```

```

; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8080
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: not relevant
; TOPOLOGY: not relevant
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
; US-09-037-601-2

Alignment Scores:
Pred. No.: 0.0252 Length: 2332
Score: 110.00 Matches: 79
Percent Similarity: 38.51% Conservative: 55
Best Local Similarity: 22.70% Mismatches: 122
Query Match: 2.44% Indels: 92
DB: 4 Gaps: 17

US-09-806-276a-4 (1-2546) x US-09-037-601-2 (1-2332)
QY 285 CCACCTTGAAGCCACTGAGAAATGATGACAGTTCCTATCAAAAAGATTACAA 344
      ||| ||| ||||| |||||
Db 749 ProserThrArgGlnLysGlnPheSnlArThrTrlPro-GlnAsnAspLLeGly 768
QY 345 A-----CATATACATCTGGAAGAAAGTGGCCCTTTC 377
      ||| ||||| ||||| |||||
Db 768 sphrAspProLrpheAlaHisArgThrProMetProLysLLeGlnAnValSerSe 788
QY 378 TCCCGCTTCAAAATATGATCATTCCAAAATGCCAGCAAGCCCATTTAC 437
      ||| ||||| ||||| |||||
Db 788 r-----AspLeuLeuMetLeuLusArgLnsr-ProThrProHisGly 803
QY 438 TGAAGCAGCAATTAACAAGAAAGAAATTAATTAAGTACAGGACATTCCTCTG 497
      ||| ||||| ||||| |||||
Db 803 euSerLeuSerAspLeuGlnLysLysLysLysLysLysLysLysLysLysLysLys 823
QY 498 ATATGATCAGT-----CCCCCGTTGGAGACTTTGCCACACATCCACA 542
      ||| ||||| ||||| |||||
Db 823 lYAlaLLeaspSerAsnAnSerLeuSerGlnMetThrHisPheArgProGlnLeuHis 842
QY 543 TTGGCAAAAGAGGCCAGACGATGCTTTGGAATATTCCTT----- 586
      ||| ||||| ||||| |||||
Db 843 -----HisSer-----GlyAspMetValPheThrProGlnSerGlyL 855
QY 587 -----CTTCAAGGAGACTGAGCTTTTAACTGGAACACAGGAGAAACACACTGGGCC 641
      ||||| ||||| ||||| |||||
Db 855 euGlnLeuArgLusAnGlnLysLysLysLysLysLysLysLysLysLysLysLys 873
QY 642 AGTTCCTGGGCAATTAAGTCTTCCGGGCCACAGCACTCCGAGCTGTGTCACAG 701
      ||||| ||||| ||||| |||||
Db 873 yslLeuAsp-----PheLysValSerSerThrSerAsnLeuLLeuSerT 888
QY 702 AAGCGCCCTCCCGGTC-----AAAATGCATCTCCCTCCCGACCATG 749
      ||||| ||||| |||||
Db 888 hrlleProseraspnleuAlaIaIaIaglyThAspAnTherSerleugly----- 905
QY 750 GAGGATCCCAAGCTCATGTTGCCCTTATGTCACAGTATTAATTCAAACAG 809
      ||| ||||| |||||
Db 906 -----ProProSerMetProValHisTyrAspSerGlnLeu 918

```



```

1  TITLE OF INVENTION: Modified Factor VIII
2  FILE REFERENCE: 75-95H
3  CURRENT APPLICATION NUMBER: US/09/315..179
4  CURRENT FILING DATE: 1999-05-20
5  EARLIER APPLICATION NUMBER: U.S. 09/037,601
6  EARLIER FILING DATE: 1998-03-10
7  EARLIER APPLICATION NUMBER: U.S. 08/670,707
8  EARLIER FILING DATE: 1996-06-26
9  EARLIER APPLICATION NUMBER: PCT/US97/11155
10 EARLIER FILING DATE: 1997-06-26
11 EARLIER APPLICATION NUMBER: PCT/US94/13200
12 EARLIER FILING DATE: 1994-11-15
13 EARLIER APPLICATION NUMBER: U.S. 08/212,133
14 EARLIER FILING DATE: 1994-03-11
15 EARLIER APPLICATION NUMBER: U.S. 07/864,004
16 EARLIER FILING DATE: 1992-04-07
17 NUMBER OF SEQ ID NOS: 40
18 SOFTWARE: PatentIn Ver. 2.0
19 SEQ ID NO 2
20 LENGTH: 2332
21 TYPE: PRT
22 ORGANISM: Homo sapiens
23 US-09-315-179-2

```

```
QY 378 TCCCGCTTCAAAATAGACATTTCAAAATCCAAATGCAAGCCAGACCCCAATTACC 437
Db 788 r-----AspLeuLeuMetLeuLeuArgGlnSer--ProThrProHisGlyL 803
QY 438 TGAAGCAGCCCAATACAGAAAGAAAGAAATTTAACTGAGGACATTCCTCTCTG 497
Db 803 euserLeuSerAspLeuIngluAlaLysTyrGluThrPheSerAspSerProSerProg 823
QY 498 ATATGATCAGT-----CCCCGCTTGAGACTTTCGCCACACCATCCACA 542
Db 823 LylAlaIleAspSerAsnSerLeuSerGluMetThrHisPheArgProGluLeuHis- 842
QY 543 TTGGCAAAAGAGGCCAGACGATGCTTTGGAGATATTTCCTTT----- 586
Db 843 -----HisSer-----GlyAspMetValPheThrProGluSerGlyL 855
QY 587 -----CTTCAAGGAGACTAGAGCTTTTACTGGAACCAAGAGAAAGACACCTGGGCC 641
Db 855 euGlnLeuArgLeuAsnGlnLysLeu-----GlyThrThrAlaAlaThrGluLeuLysL 873
QY 642 AGTTCCCTGGGCAATATGAGTTCTTCCGGCCCAACAGCCTCGACTGTGTTCAAG 701
Db 873 LysLeuAsp-----PheLysValSerSerThrSerAsnAsnLeuIleSerT 888
QY 702 AAAGCCCTTCCCGGCTCTC-----AAAAATGCCATCTCCTCCGACCATTG 749
Db 888 hrIleProSerAspAsnLeuAlaAlaGlyThrAspAsnThrSerSerLeuGly----- 905
QY 750 GAGATCCCAAGCTCATGTCCTTATGTCACCAAGATTAATTCACAAACAG 809
Db 906 -----ProProSerMetProValHisTyrAspSerGlnLeuA 918
QY 810 AGTCC-----TTGGGGCCAGCAAGCTGCCAGGCTT--AGCTGCGAGCCGTCATGG 860
Db 918 spThrThrLeuPheGlyLysSerSerProLeuThrGluSerGlyLysProLeuSerL 938
QY 861 AGAAAGAGCTCAGAGAAAGAGAGTCTGTGAGAGATGGACAGTCCACAGGAGACA 920
Db 938 euserGluGluAsnAsnAspSerLysLeuLeuGlnSerGlyLeuMetAsnSerGlnGln 958
QY 921 CCTGTGGGGCTCCAGCGCTTTCATCTCAGTCCAGC-----CAAGGAGAGACA 971
Db 958 erSerTrpGlyLysAsnValSerSerThrGluSerGlyArgLeuPheLysGlyLysArg 978
QY 972 GCCACTCTCCAGCCTGTCCGACAGTACCCTGCGCAGCGCAGGACATGTTTGACC 1031
Db 978 lahIsGlyProAlaLeuLeu-----ThrLysAspAsnAlaLeuP 991
QY 1032 ATCCACCCCATGCGAGCTCATCAAGAGAAAGACTAAGTACAGAGAGTCCCTCTGAC- 1090
Db 991 helysValSerIleSerLeuLeuLysThrAsnLysThrSerAsnAsnSerAlaThrAsnA 1011
QY 1091 -----CTTACAGGTTCCCTCTCTCCCTGACAGTTCGAGCTTGAGCCCTCACTT 1139
Db 1011 rglYsThrHisIleAspGlyPro-----SerLeuLeuIleGluAsnSerProSerValT 1029
QY 1140 TGGATGAGGTGCTGAAT----- 1156
Db 1029 rpgLAsnGlnLeuGlnLeuSerAspThrGlnPheLysValThrProLeuIleHisAspA 1049
QY 1157 -----GTAATGGATATAAAT 1171
Db 1049 rgmEtLeuMetAspLysAsn 1055
```

Search completed: July 18, 2003, 17:40:43  
Job time : 81.9421 secs







## RESULT 3

US-10-106-698-4753  
 ; Sequence 4753, Application US/10106698  
 ; Publication No. US20030109690A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ruden et al.  
 ; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
 ; FILE REFERENCE: PA005P1  
 ; CURRENT APPLICATION NUMBER: US/10/106,698  
 ; CURRENT FILING DATE: 2002-03-27  
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524  
 ; PRIOR FILING DATE: 2000-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/157,137  
 ; PRIOR FILING DATE: 1999-09-29  
 ; PRIOR APPLICATION NUMBER: US 60/163,280  
 ; PRIOR FILING DATE: 1999-11-03  
 ; NUMBER OF SEQ ID NOS: 8564  
 ; SOFTWARE: PatentIn Ver. 3.0  
 ; SEQ ID NO 4753  
 ; LENGTH: 149  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: MISC\_FEATURE  
 ; LOCATION: (85)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: MISC\_FEATURE  
 ; LOCATION: (136)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: MISC\_FEATURE  
 ; LOCATION: (145)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; US-10-106-698-4753

## Alignment Scores:

Pred. No.: 0.000291 Length: 149  
 Score: 148.00 Matches: 61  
 Percent Similarity: 35.75% Conservative: 13  
 Best Local Similarity: 29.47% Mismatches: 46  
 Query Match: 3.29% Indels: 87  
 DB: 15 Gaps: 12

US-09-806-276a-4 (1-2546) x US-10-106-698-4753 (1-149)

QY 395 ACATTCCTCAAAATTCGAAATGCGAGCCAGCCCAATTTAC----- 436  
 Db 1 SerTprSerArgGluArgAlaProAla-----ProLeuTprGluAspArgGluMetPro 18  
 QY 437 ---CTGAAAGCAGCCCAATTAACAAGAAAGAAATTTAACTGAGGCGCATTCGTCT 493  
 Db 19 ValLeuLysGlnLeuGlyProAlaGlnProLysLysArgProAspArgGlyAlaLeuSer 38  
 QY 494 CCTGATATGATCAGTCCCGCTGAGACTTCGCGACCATTCACATTTGGCAAAGAG 553  
 Db 39 -----IleSerAlaProLeuGlyAspPheArgIsthrLeuHisValGlyArgGly 55  
 QY 554 GCGCCAGCAGCATGCTTTTGGAGATATTTCTTTCTTCAAGGAGACTACGACTTTTACT 613  
 Db 56 Gly-----AspAlaPheGlyAspThrSerPheLeu----- 65  
 QY 614 GGAAGACGAGAGAGACACACTGGGCGCACTTCCTCGGCATATAGAGTTCTTCGGGGCC 673  
 Db 66 -----SerArgHisGlyGlyGly----- 71  
 QY 674 AACAGACACTTGAGCTGTGTTCACAGAAAGCCCTCCCGGTCTGCAAAAATGCCATC 733  
 Db 72 -----ProPro-----Prose 75  
 QY 734 TCCCTCCGAGCATTTGGAGAGTCCCAAGCTCATAGTTGCCCTATTGTTCACAGTGACA 793  
 Db 75 rProGlyArgProArgGlyProArg----- 84

QY 794 TTATATCCAAACAGAGTCTTCGGGCGCAGAAAGCTGCCAGGCTTA---GCTGGCAG 850  
 Db 85 -----ProArgArgArgArgArgProGlnSerAlaProAlaGluLeuArgProAlaVala 103  
 QY 851 CCGCTCA-----TGAGGAGAAAAGCTCAGAGAGAAAAGCAGTCTG 889  
 Db 103 lProSerProGlySerGlyAlaSerCysTrpThrArg-----Cy 116  
 QY 890 TTGAGAGATGGAGACATCCAGCAG-----GAGACACTCGCTG 928  
 Db 116 sTrpArgMetAspAlaAlaArgArgSerGlyCysAlaSerHisAlaAsnProProGly\*\* 136  
 QY 929 GCGTCCAGCGGTTCGTCAT 947  
 Db 136 \*AlaProAlaValaValaGlnHis 142

## RESULT 4

US-10-156-761-10907  
 ; Sequence 10907, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OKURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 10907  
 ; LENGTH: 797  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 ; US-10-156-761-10907

## Alignment Scores:

Pred. No.: 0.019 Length: 797  
 Score: 130.00 Matches: 97  
 Percent Similarity: 33.50% Conservative: 36  
 Best Local Similarity: 23.82% Mismatches: 167  
 Query Match: 2.89% Indels: 101  
 DB: 15 Gaps: 14

US-09-806-276a-4 (1-2546) x US-10-156-761-10907 (1-797)

QY 5 GCTAGCCCGGAGAACCGCGGACCGGCTGGGGCGCTTCAAGCGCGTCTC---GAGAC 58  
 Db 393 SerAlaProSerProLeuProProAlaSerGlyAla-GlnArgSerValIleCysGlyAl 412  
 QY 59 GGAATATGGCGGAGCAGAG----- 77  
 Db 412 aserValAlaThrAlaThrAlaArgArgThrGlnLysGlyPheSerValSerSerThrArgVa 432  
 QY 78 -----CACACGCGCGGACAGATCGGCTGGATCGTGGACAGCGGAGGAGTGGCTGC 130  
 Db 432 lThrLeuArgArgAsnGlnArgLeuValTrpAlaLeuAlaGlnArgSerHisValProCy 452  
 QY 131 GCCCGGAGAGCGCGGCGGCGGCGGCTGGATCTTCAAGCGGCGCGGTTTCTCTGTT 190  
 Db 452 s---AlaGlyCysProGluSerArgLeu----- 460  
 QY 191 GTGTCAAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 250  
 Db 461 -----ArgAlaLeuAlaGProLeuProSerPro----- 469

```

QY 251 GGAAGTCTCTGTTGACCCCATCTCATTTGGAAGCCACTTTCAGACCCACTTGAGAAA 310
    |||||
Db 470 -----SerineProProAlaAlaThrArgProAlaSerProArgProAlaAla 487
QY 311 TGATGTGACAGTTCCTATCAAAAAGATTCAGAAACATATACCATCTGTGAAGAGTG 370
    |||||
Db 487 aAlaAlaAlaArgProAlaAlaLeuProSerProThrThrSerAlaAlaAlaArgAlaThrSe 507
QY 371 CCCCTTTCGCCGCTTCGCAAAATAGACATTCCTCAAAATTCGCAAAATGCCAGCAAGCCCA 430
    |||||
Db 507 rProSerThrThrLeuProProProAlaThrArgArgProAlaArgSerArgProPro 527
QY 431 ATTTCCTGGAAGCAGCCCAATACAGAAAGAAAGAAATTAAGTACGAGGACATTCTG 490
    |||||
Db 527 o-----ProArgProSerSerArgArgThrAlaSer----- 537
QY 491 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 550
    |||||
Db 538 -----ProThrProThrArgGlySerValTrpSerSerSerProSe 551
QY 551 GAGGCCAGCAGAGTCTTTGAGATATTCCTTCTTCAAGGAAGTACGAGCTTTTA 610
    |||||
Db 551 rProAlaThrThrArgSerValSerAlaSerProThrProProSerArgLysProPr 571
QY 611 CCGGAAGACGAGAGAAACACACACCTGGCCAGTTCCCTGGCATATAGTTCTTCGG 670
    |||||
Db 571 oArgSerThrArgArgProProSerAlaSerSerThr-----ThrSerArgSerAr 588
QY 671 GCCACAGCAGCCTGACCTGTGTTCACAGAAAGCCCTCCCGGTGCTCAAAATCC 730
    |||||
Db 588 gArgThrThrSerArgThrSerSerSerThrArgArgArgProProThrSerProAlaSe 608
QY 731 ATCTCCCTCCGACCATTTGGAGATGCCAAGCTCTCATGTGCTTATTTGTCACAGT 790
    |||||
Db 608 rProProArgPro----- 613
QY 791 ACATTTAATTCACAAAGAGTCTTCGGCCAGCAAGCTGCCAGCTTAGCTGCAG 850
    |||||
Db 614 -----ProSerArgThrSerSerAlaSerSerAlaAlaThrThrSerArgSerTh 630
QY 851 CCGGTCATGAGGAAAGAAAGCTTCAGAGAAAGCAAGTCTGTGGAGATGGGACAGTCC 910
    |||||
Db 630 rSerSerArgArgAlaSerArgAlaSerArgThrProArgArgValSerArgSerSe 650
QY 911 CAGGAGACACCTCGTGGGCTCCAGCGGTTTCGACATCTGACCCAGCCAGCAGAGAC 970
    |||||
Db 650 rArgSerThrProArgProProProAlaAla-----GlySerProAlaAlaArgThrArgAl 668
QY 971 AGCCACCTCTCCA-----GCCGTTCGAGACAGTACCCCGAGCTGCCAGCCGAGGAC 1021
    |||||
Db 668 aArgProArgProAlaAlaArgSerAlaProArgProThrSerThrProProAlaAl 688
QY 1022 ATGTTTGACATCCACCCCATGGAGCTCA-----TCAAGGAGAAAGACT 1066
    |||||
Db 688 acys-----ProAlaAlaAlaSerSerArgProProArgProArgThrArgGlySargAr 705
QY 1067 AAGTCAGAGAGTCCCTCTGACC-----TTACAGGTTCCCTC 1105
    |||||
Db 705 gSerThrProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 725
QY 1106 CTCTGCC 1112
    |||||
Db 725 rSerPro 727

```

## RESULT 5

US-10-260-715-8  
 : Sequence 8, Application US/10260715  
 : Publication No. US2003009992A1

## GENERAL INFORMATION:

: APPLICANT: UCB, S.A.  
 : APPLICANT: NOCKA, Karl  
 : APPLICANT: LU, Sun  
 : APPLICANT: MEDLEY, Quintus

```

: APPLICANT: THOMIS, Daniel
: APPLICANT: GU, Jessie
: TITLE OF INVENTION: Genes Associated with Mast Cell Activation
: FILE REFERENCE: 053529-5006
: CURRENT APPLICATION NUMBER: US/10/260,715
: CURRENT FILING DATE: 2002-10-01
: PRIOR APPLICATION NUMBER: 60/325,536
: PRIORITY FILING DATE: 2001-10-01
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 8
: LENGTH: 1259
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-10-260-715-8

```

```

Alignment Scores:
Pred. No.: 0.0669 Length: 1259
Score: 124.50 Matches: 101
Percent Similarity: 29.15% Conservative: 29
Best Local Similarity: 22.65% Mismatches: 178
Query Match: 2.77% Indels: 138
DB: 15 Gaps: 18

```

US-09-806-276A-4 (1-2546) x US-10-260-715-8 (1-1259)

```

QY 6 CTAGCCGAGAGACCCGCGCAGCCCTGCGGCGCTTCACGCGCTTCGAGCGGATAT 65
    |||||
Db 282 LeuAlaGluProAla-----LeuProAlaLeuProAlaProAlaLeuGlnProGly 300
QY 66 GCGGTGAGCAGCAGCAGCCGCGGAGACTGGCTGTGATCTGCGCAGCGGAGATG 125
    |||||
Db 301 SerGlnThrGluGlySerGlyAlaLysGlyLysP-----SerArgGluAlaSerGlyVal 319
QY 126 CGTGGCGCGCGGAGCGCGCGGCGGAGCGGCTGGATCCTCAGCGCGCGGTTTGCC 185
    |||||
Db 320 ProAlaProGlyLysLysLys----- 325
QY 186 TGGTTGTG--TCAAGACTGATGATGTAAGTGGCTCTCTGAGAGCTTCAGTGGCG 242
    |||||
Db 326 TrpProTrpValSerArgGluValProGlyThrArgSerPheGly----- 340
QY 243 TAACCTCAGAGAGTCTCTTTCAGCCCATCTCATTTTCAGAACCACTTCTGAGCCATT 302
    |||||
Db 341 ---ProAlaProGlySerThrArgProTrpLeuGlnSerProProGlnGlyArgProLeu 359
QY 303 GAGAAAT-----GATGTACAGTTCCTATCAAAAGAT 338
    |||||
Db 360 SerSerGlnGlyProGlyAlaThrGlyAlaTyraSpAlaGlyGluAlaLysPser 379
QY 339 TCAGAACATATACCATCTGTGAAGAGTGGCCCT----- 374
    |||||
Db 380 SerArgSpAsnSerProAlaAlaSpLeuGlyProThrArgProProGlnGlnAlaLys 399
QY 375 -----TCTCTCCGCTTCGCAAAATAGA-----CATTCT 401
    |||||
Db 400 ProAlaAlaAlaGlyHisSerSerArgAlaProSerThrSerArgLysLysLysLysLys 419
QY 402 CAAATTCGCAAAATGCGCAGCAGCCCAATTTACTGAAAGCAGCCATATACAGAAAG 461
    |||||
Db 420 AlaSerProProAlaAlaAlaProGlyProGlyPhePro----- 431
QY 462 GAAAGAAATTTAAACTGAGGACATTCTGTCTCTGATGATGATGATGATGATGATGATG 521
    |||||
Db 432 -----ProGluAlaLeu 435
QY 522 ACTTTCGCGACACCATTCACATTTGCGAAGAGGCGCAGCAGATG----- 566
    |||||
Db 436 ThrLeuProSerProSerAspPheLeuProLeuGlnValThrGlnAspProSerValGly 455
QY 567 ---TCTTGGAGATATTCCTTCTTTCAGAGGAACACTGAGCTTTTACCTCGAAGACGAG 623
    |||||
Db 456 GluAsnLeuArgAlaAlaProAlaProAlaProSerSerAlaSerAlaGlnValLeuThrSerAla 475

```

```

QY 624 AGAAGACACCTGGCCATTCCTGGCATATGATGTTCTCCGGGCAACAGCACT 683
Db 476 ProAlaSerValLeuAlaProAlaLeuAlaSerSerProSerSerAlaProThSerAla 495
QY 684 CGGACTCTGTGTTCACAG-----AAAGCCCTCCCGG 716
Db 496 ThrThrSerThrSerSerProThSerAlaProAlaProThSerAlaProThr 515
QY 717 TGCTCAAAATGCAATCCCTCC----- 740
Db 516 SerThProAlaProAlaProSerProAlaAlaAlaThProAlaProAlaProAl 535
QY 741 -----CGACCATGGAGATCCCAAGCTC 764
Db 536 ProValProThLeuThProProSerProAlaLeuThProValProThProAlaLeu 555
QY 765 TCATGTTGCCCTTATGTACACAGACATTAATTCGCAACAGAGAGCTCTGGGGCCAG 824
Db 556 SerProAlaPro-----ThProAlaProThProAlaAlaSer 568
QY 825 CAAGC-----TGCCAGGCTTACGTGCGAGCCCTCATGAGAAAGAGCTCAGG 875
Db 569 ProAlaLeuThProValProThProAlaLeuSerProAlaProThProAlaLeuThr 588
QY 876 AGAAAGCAAGTCTGTGGAGATGGACAGTCCACAGGAGACACCTCTGGGGCTCCA 935
Db 589 ProAlaAla-----SerProAlaLeuThProValProThPro 601
QY 936 GCGGTTCTGATCTCACTCCAGCCAGGACAGACAGCAGCTCTCCAGCTCCGAG 995
Db 602 Ala-----LeuSerProAlaProThProAlaProThProAlaAlaSerProAla 618
QY 996 AGTACCCGACGTGGCCAGCCAGG---ACATGTTTGACATCCCAAGCCCATGCAAGCTCA 1052
Db 619 ProAlaProThSerAlaProThProThProAlaAlaSerProAlaProAlaSpoly 638
QY 1053 TCAAGGAAAGACTAGTCAAGAGAGTCCCTCTGACCTTACAGGTCTCCCTCTCC 1112
Db 639 SerLys-----ProIngluSerValAlaLeu-ProArgLysIrgIngluGly 655
QY 1113 TGCAAGCTTCACTCTGG 1128
Db 655 nValSerAlaSerTyr 660

RESULT 6
US-10-086-464-5
; Sequence 5, Application US/10086464
; Publication No. US2002019218A1
; GENERAL INFORMATION:
; APPLICANT: GORING, Daphne R. et al.
; TITLE OF INVENTION: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES
; FILE REFERENCE: P 25,762-A USA
; CURRENT APPLICATION NUMBER: US/10/086,464
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 10/069,304
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: PCT/CA00/00966
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: US 60/149,466
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: US 60/159,122
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 721
; TYPE: PRT
; ORGANISM: Brassica napus
US-10-086-464-5
Alignment Scores:
Pred. No.: 0.061 Length: 721

```

```

Score:
Percent Similarity: 124.00
Best Local Similarity: 34.76%
Query Match: 27.62%
DB: 15
Gaps: 9

US-09-806-276a-4 (1-2546) x US-10-086-464-5 (1-721)
QY 491 TCTCCGATATGATCACTCCCGCTGGAGACTTCCGACACCATTCACATGGCAAA 550
Db 5 SerProCysLeuLeuSerSerProAlaAlaSerPhe--HisSerLeuHis-Leu----- 21
QY 551 GAGGCGCAGCAGCATGCTTTGGAGATATTCCTTTCTTCAAGGGAAGTACAGACTTTTA 610
Db 22 -----ProPhePhe----- 24
QY 611 CTGGAAGAACAGAGAACACACCTGGCGCAGTTCCTGGGCATATGACTTCTCCG 670
Db 25 -----ArgSerLysGluAlaAspMetSerAlaProSerProGlyThrGlyse 41
QY 671 GCCAAGACAGCCTCGAGACTCTGTTCACAGAAAGCCCTCCCGGTCAAAATGCC 730
Db 41 rProProSerProProSerAsnSerThrThrThrProProProAlaSer--AlaPr 60
QY 731 ATCTCCCTCCGACCATTTGAGAGATCCCAAGCTCTCATGTTGCCCTTATGTCACGAG 790
Db 60 oProProThrThrProSerSerProProProProSer----- 72
QY 791 ACATTTAATTCGCAACAGAGATCTTGGGCGCAGAAAGCTGCCAGGCTAGCTCGAG 850
Db 73 -----ThrIleProThrSerProProProSerSerArgSerThrProSer--AlaProPr 90
QY 851 CCCGTCATGAGGAGAAAGCTCAGAGAAAGAGCTGTGTGAGATGGACAGTCCAC 910
Db 90 oProSer-----ProProTh 95
QY 911 CAGGAGACACCTCTGTGGGCTCCAGCGGTTCTGCATCTCACTCCAGCAGCAAGAGAG 970
Db 95 rProSerThrProGlySerProProProLeuProInProSerProProAlaProThrTh 115
QY 971 AGCCACTCTCTCAGAGCTGTCCGACAGTACCCGAGTGGCCGCGAGACATGTTGAC 1030
Db 115 rProGlySerProProAlaProValThrProThrThrAsnProProProSerValPr 135
QY 1031 CATCCACCCCATGAGCTCATGAAGGAAAGACTAGTCAAGAGAGTCCCTCTGAC 1090
Db 135 oGlyProPro-----SerAsnProSerArgIrgIngluGlySerProArgProProSer----- 152
QY 1091 CTTCAGAGTTCCTCTCTCCCTGAGC 1118
Db 153 -SerProSerProProSerProSerSer 161

RESULT 7
US-10-128-714-3298
; Sequence 3298, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Ulang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshekin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890

```



QY 702 AAAGCCCTCCCGGCTTCATAAAATGCCATCTCCCTCCGACATTGGAGATCCCAAG 761  
 DB 512 LysProProProArgSerSerSerIleProAlaPro-----ProLeuLysLaserTh 529  
 QY 762 CTCACATGTCGCCCTATTGTCTACACAGTACATTTAAATCCAAACAGAGTCTTCGGGC 821  
 DB 530 ValSer-----ProAlaProLeuLys 536  
 QY 822 CACCAAGCTCCGACGCTTACCTGCGAGCCGCTCATGAGCAAAACCTC----- 872  
 DB 537 ThSerSerThAlaProThrAlaProGlnLysAsnPheGluGluLeuProLeuPro 556  
 QY 873 -----AGAGAAAGCACTCTGTGGAGAAATGGACAGATCCACCGAGGAGACACTCGT 926  
 DB 557 ProProArgProLysSer-----ArgProLaserSerGlyArgTyrThrPro--- 572  
 QY 927 GGGGCTCCAGCGTTCTGATCTCAGTCAGCCAGCAAGAGACAGCAGCCACTCTCCAGCC 986  
 DB 573 AsnAlaProVal-----SerAlaProSerLeuProGlnSerIleProProAla 589  
 QY 987 TGTCCGACAGTACCCCGACCTGCGACGCGAGACATGTTGACCATCCCA---CCCCAT 1043  
 DB 590 AsnGlnTyrSerAsnValProGlnAlaProGlnSerAsnIleGlyProProAspProPro 609  
 QY 1044 GCGAGCTCATCAAGGAAAGACTAGTACAGAGAGTCCCTCTCTGACCTTACAGAGTTCC 1103  
 DB 610 GlnLeuGlnGlnProGlnLysArgLeuAspProTyrSerAsnLeuAlaProAsnValPro 629  
 QY 1104 TCCTCTCCC 1112  
 DB 630 SerAlaPro 632

## RESULT 9

US-09-729-674-158  
 : Sequence 158, Application US/09729674  
 : Patent No. US2001003935A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Jacobs, Kenneth  
 : APPLICANT: McCoy, John M.  
 : APPLICANT: Lavallee, Edward R.  
 : APPLICANT: Collins-Racie, Lisa A.  
 : APPLICANT: Evans, Cheryl  
 : APPLICANT: Treacy, David  
 : APPLICANT: Meiberg, Maurice  
 : APPLICANT: Agostino, Michael J.  
 : APPLICANT: Steininger II, Robert J.  
 : APPLICANT: Spaulding, Vikki  
 : APPLICANT: Wong, Gordon G.  
 : APPLICANT: Clark, Hilary  
 : APPLICANT: Fechtel, Kim  
 : APPLICANT: Genetics Institute, Inc.  
 : TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
 : FILE REFERENCE: 6055-64X  
 : CURRENT APPLICATION NUMBER: US/09/729,674  
 : PRIOR FILING DATE: 2000-12-04  
 : PRIOR APPLICATION NUMBER: 09/539,330  
 : NUMBER OF SEQ ID NOS: 283  
 : SOFTWARE: PatentIn Ver. 2.0  
 : SEQ ID NO 158  
 : LENGTH: 586  
 : TYPE: PRT  
 : ORGANISM: Homo sapiens  
 : FEATURE:  
 : NAME/KEY: UNSURE  
 : LOCATION: (286)  
 : US-09-729-674-158

Alignment Scores:  
 Pred. No.: 0.42  
 Score: 114.00  
 Percent Similarity: 32.74%

Length: 586  
 Matches: 79  
 Conservative: 32

Best Local Similarity: 23.30%  
 Query Match: 2.53%  
 DB: 10  
 Gaps: 19

US-09-806-276a-4 (1-2546) x US-09-729-674-158 (1-586)

QY 306 AAAAATGATGTGACACTTCTATCAAAAAGCATCA-----GAAACATATACCATCTGT 359  
 DB 243 LysAsnThrVal-----LysAspSerIleAlaGlnLaserSerValLys 257  
 QY 360 GAA-----GAAATGGCCCTTTCTCCGCTTGGCAAAATAGACATTTCAAATTCCA 410  
 DB 258 GluLeuSerMetAspSerAsnGlnLysPheSerLys---GluGlyAlaAlaGluProGluPro 276  
 QY 411 AATGGCCAGCCCAAGCCCAATTTCTGAAAGCAGCAGCAATTAACAAAGAAAGAAAT 470  
 DB 277 GlnSerLeuPro-ProSerAsnLeuSer\*\*\*SerSerGlnGlnProAspIleAsnLe 296  
 QY 471 T-----TAAACTGAGGACATTCTGTC---TCCTGATATGA 503  
 DB 296 LysPleuSerProLeuThrIleuGlySerProGlnAsnHisThrLeuGlnAlaProGlyG 316  
 QY 504 TCAGTCCCGCCGCTTGG---AGACTTTCGCCACACCATCCACATTGGCAA-----AGAG 554  
 DB 316 LysAlaProProThrAlaGluMetArgGlyProHisProProTyrThrGluValArgG 336  
 QY 555 GCCA----- 558  
 DB 336 LysProProGlyLysValProGluAsnGlyLeuValArgArgLeuAsnThrValProAs 356  
 QY 559 -----GCACAGTCTCTTGGAGATATTTCCTTTCTTCAAGGAAACTAGACCTTTTACCTG 614  
 DB 356 LysSerArgValIleThr-----ValLysThrPr 366  
 QY 615 GAAACACAGA----- 626  
 DB 366 LysProGlyLysProSerSerGlnLysProSerSerLysGluValProSerCysLysG 386  
 QY 627 AAGCACACCTGGGCACTTCCCTG-----GGCAT 655  
 DB 386 GlnLeuProGlnProValSerSerLysGlyLysProGlnLysGlyLysArgGlnLysG 406  
 QY 656 AATGATTTCTCCGGGCAACAGCACCTCGACTGTGTCAAGAAAGCCCTCCCG 715  
 DB 406 GlnAlaLysLysSerGlnLaserPro-----AlaProAlaGlyProAla 421  
 QY 716 GTGCTCAAAATGCCATCT-----CCCTCCGACG 745  
 DB 421 AsnLeuGluValProSerAlaLysGlnValAlaGlyProLysGlnProGlyArgG 441  
 QY 746 ATTGAGAGATCCCAAGCTCTCATGTTGGCCCTTATTGTCAACAGTACATTTAATTCAAA 805  
 DB 441 LysGlnLeuProLysValGlySerCysAlaGlnAla-----G 454  
 QY 806 CAGGAGTCTTCGGGCAACAAAGCTGCCCAGCTTACGTGGAGCCCGTCATGGAGAGAA 865  
 DB 454 LysGlnLysArgLysArgProGlyProGlyThrAlaGlySerProLysThrGlnLys 474  
 QY 866 AAGCTCAGAGAGAAAGCAGTGTGTGGAGAAATGGAGATGCCACAGAGACAGACCTCG 925  
 DB 474 SerGlu-----LysLysSerSerTrpArgAsnTrpProG 485  
 QY 926 TGGGCTCCAGCGGTTTGCATCTCAGTCCAGCCAAAGCAGACAGACACTGCT 980  
 DB 485 LysAlaLysAla-----ArgProGlnGlnGlnLysSerValGlnProProG 501  
 QY 981 -CCAGCTGTCCGACAGTACCCGACCTGGCCAGCCAGACATGTTGACCATCCACCC 1039  
 DB 501 LysAlaArgProGlnSerLeuProGlnGlnLysGlyArg----- 514  
 QY 1040 CCATGCGAGCTCATCAAGGAAAGACTAAGTCAGAGAGTCCCTGCTGACCTTA 1094  
 DB 515 -----SerArgLysSerArgAsnLysGlnGlnLysProAlaSerSerLeu 529



US-10-156-761-14622

Alignment Scores:

Score:

Best Local Similarity

DB:

US-09-806-276A-4

Q. 10

22

1

1  
2  
3  
4  
5

43 11e1

128 TGC

65 -Ala

QY 188 GTT

83 p. --- Db

248 TCAG

— 1 —

2007-2008

• •  
• •  
• •  
• •

1

33

0270 0000

411

148 aArgi

452 AACAF

168 PTH

512 correct

三

[REDACTED]

```

Db      173  oargtrpArgThrSerProLysProAlaGlySerThrValArgGlyTyrPseThrThrAr 193
Oy      555  -----GCCAGCAGATGCTCTTGGAGATATTCTCTTTCAGAGAACTAC 601
Db      193  gThrGlySerAlaSerSerAlaThrAlaAlaGlyAlaProProGlyArgArgGly----- 211
Oy      602  GAGCTTTACCTGGAACAGAGAGACACACCTGGCCAGTTCCTCGCATATGAG 661
Db      212  -----SerArgAlaArgArgThrTyrPseArgSerArgCysProSerAr 226
Oy      662  TTCCTCCGGCCACACAGCAGCTCGACTCTGTCTGTCACAGAA----- 704
Db      226  gSerAlaAlaArgAlaAlaPro-----CysSerValArgProGlySerThrThrse 243
Oy      705  -----GGCCCTCCCGGCTGCTCAAAATGCCATCT--CCCTCCGACCATGGAGATCC 757
Db      243  rThrArgProValArgSerSerSerThrProGlyTyrProArgArgArgArgAlaArgp 263
Oy      758  C-----AAGCTCTCATGTGGCCCTTATTTGACACAGTACATTATTCACAAAGAG 811
Db      263  oSerArgArgCysGlyCysCys-----AlaAlaAsnArgse 275
Oy      812  T---CCTGGGGCCCAAGCTGCCAGCTTACCTGCGAGCCCGTCATGAGAGAA 868
Db      275  rValProHisGly-----ProValArgTyrPargArgas 286
Oy      869  GCTC-----AGAGAAAGACACTGTGTGGAGATGGAGACAGTCCACAGGAGAC 919
Db      286  nCysThrProGlyArgArgArgAlaThr--LysArgTyrGlyArgThrProPse 305
Oy      920  ACCTCGTGGGGCTCCA-----GGGTTCTGCA 946
Db      305  rProCysGlyArgProProProSerArgArgCysGlyArgArgThrAlaGlyArgAr 325
Oy      947  TCTCAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1006
Db      325  gValSerProSerAlaValSerAlaHisArgPro-----CysProThrAlaSerProG 343
Oy      1007  TGGCCAGCCGAGAGCATGTTGACCATCCACCCCATGCGAGCTCATCAAGGAAAGCT 1066
Db      343  yGlyPro-----Le 346
Oy      1067  AAGTCAGAGAGATCCCTCTCTGACCTTACAGTTCC----- 1103
Db      346  uAlaArgThrSerProThrValArgSerThrValProCysGlySerAlaArgArgLyn 366
Oy      1104  -----TCCTCTCCCTGC 1115
Db      366  sThrSerAlaArgProCys 372

RESULT 13
US-10-142-515-11
: Sequence 11, Application US/10142515
: Publication No. US20030078399A1
GENERAL INFORMATION:
: APPLICANT: SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH
: APPLICANT: Lloyd, Kenneth O.
: APPLICANT: Ylin, Beatrice W.T.
: TITLE OF INVENTION: Nucleic Acid Sequence Encoding Ovarian Antigen, CA125, and Uses T
: FILE REFERENCE: 649-A-US
: CURRENT APPLICATION NUMBER: US/10/142.515
: CURRENT FILING DATE: 2002-07-23
: PRIOR APPLICATION NUMBER: US 60/290,480
: PRIOR FILING DATE: 2001-05-11
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 11
: LENGTH: 5877
: TYPE: PRT
: ORGANISM: Human Being
: FEATURE:
: NAME/KEY: MISC_FEATURE
: LOCATION: (1)..(5877)

```

OTHER INFORMATION: Amino acid sequence of MUC16B  
US-10-142-515-11

```

Alignment Scores:
Pred. No.: 1.53
Score: 111.50
Percent Similarity: 32.49%
Best Local Similarity: 22.01%
Query Match: 2.48%
DB: 15
Gaps: 21

US-09-806-276a-4 (1-2546) x US-10-142-515-11 (1-5877)
Oy      33  GGGGGCCCTTCAGCGCGTCTCGAGCGATATGCGGTAGCAGCAGCAGCAGCAGCAG 92
Db      350  GlyMetProThr-----GlyAlaThrThrLysValSerArgThrGluAlaLeuSer 366
Oy      93  CTCGGCTGATCTGCGCAGCAGCGGAGGATGCTGGCGCCGGGAGCGCGGCGAG 152
Db      367  LeuGly-----ArgThrSerThrProGlyProAlaGln 377
Oy      153  CGG-----CTGGGA 161
Db      378  SerThrIleSerProGluIleSerThrGluThrIleThrArgIleSerThrProLeuThr 397
Oy      162  TCTCTAGCGCGCGCGGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 221
Db      398  ThrThrGlySerAlaGluMetThr--IleThrProLysThrGlyHisSerGlyAlaSer 417
Oy      222  CTAGAAACCTCAGCTTGGCGCTACCTCAGAAAGTTCCTTTGACCCCATCTCATTTG 281
Db      417  eGlnGlyThrPheThrLeuAspThrSerSerArgAlaSerThrProGlyThrHisSer 437
Oy      282  AAGCCACTTCTGAAAGCAGCTTGAAGAAATGATGTGACAGTTCCTTCAAAAGATTC 341
Db      437  lAlaThrHisArgSerPro--HisSerGlyMetThrThrPrometSerArgGlyProg 456
Oy      342  GAAACATATACCATCTGTGAAGAAAGTGGCCCTTCTCCGCTTGCAGAAATAGAC 397
Db      456  lAspVal-----SerThrProSerArgProSerValGluLysThrSerP 471
Oy      398  -----TTCTCAATTCGCAAAATGCCAGCAGCAGCAGCAGCAGCAGCAGCAG 443
Db      471  roProSerSerLeuValSerLeuSerAlaValThrSerProSerProLeuTyrSerThrP 491
Oy      444  CAGCAGATATACAGAGAGAGAAAGAAATTAACCTGAGGACATTCCTCCCT 496
Db      491  roSerGluSerSerHisSerSerProLeuArgValThrSerLeuPheThrProValMet 511
Oy      497  -----GATATGATC-----AGTCCCGC----- 514
Db      511  eLysThrThrAspMetLeuAspThrSerLeuGluProValThrThrSerProSer 531
Oy      515  -----CTTGAGACTTGGCCACAC-----A 536
Db      531  etaenIleThrSerAspGluSerLeuAlaThrSerLysAlaThrMetGluThrGluAla 551
Oy      537  TCCACATTTGCAAGAGAGCGCAGCAGCAGATCTTGGAGATTTTCTTCTTCAGGGA 596
Db      551  leGlnLeuSerGluAsnThrAlaValThrHisMetGlyThrIleSerAlaArgGlnGlu 571
Oy      597  ACTAGACCTTTTACCTGGAACACAGAGAGAGAGACACACTGGGCCAG----- 643
Db      571  heTyrSerSerTyrProGlyLeuProGluProSerLysValThrSerPrometValThr 591
Oy      644  -----TTCCCTGGCAGTAATGAGTTCTTCGGG 671
Db      591  erSerThrIleLysAspIleValSerThrThrIleProAlaSerSerGluIleThrArg 611
Oy      672  CCACAGCAGACCTCGAGCTGTGTTCACAGAAAG----- 706
Db      611  leGluMetGluSerThrSerThrLeuThrProThrProArgGluThrSerThrSerGln 631

```



QY 707 -----CCCTCCCGGTGCTCAAAATGCGCATCTCCCTCCGACCA 746  
 DB 631 LuIleHisSerAlaThrLysProSerThrValProTyrLysAlaLeuThrSerAlaThrI 651  
 QY 747 TTGGAGATGCCCAAGCTCTCATGTGGCCCTTATGTCACAGATTAATTCACAAAC 806  
 DB 651 legIuaspSerMetThrGlnValMet-----S 660  
 QY 807 AGGAGTCCTTCGGGCGCAGCAAG-----829  
 DB 660 erSerSerArgGlyProSerProaspInSerThrMetSerGlnaspIleSerThrGluV 680  
 QY 830 --CTGCCAGGCTAGTGGAGCGGCTCATGGAGAAAAGCTCAGAGAGAAAAGCAGTC 887  
 DB 680 alIleThrArgLeuSerThrSerProIle-----LysThrGlnSerThrGluMet 697  
 QY 888 TGTGGAGAAATGGGACAGTCCACACGAGAGACACCTGTTGGGCTCC-----934  
 DB 697 hrIleThrThrGlnThrGlySerProGlyAlaThrSerArgGlyThrLeuThrLeuasp 717  
 QY 935 -----AGCGTTCTCATCTCAGTCCAGCAGGACAGAGACAGCCACT 977  
 DB 717 hrSerThrThrPheMetSerGlyThrHisSerThrAlaSerGlnGly---PheSerHis 736  
 QY 978 CC-----TCCAGGCTGTGCGAAGACAGTACCC-----GACTGGCCAGCCGAGAGACA 1022  
 DB 736 erGlnMetThrAlaLeuMetSerArgThrProGlyGluValProTyr-----751  
 QY 1023 TGTGTGACCATCCAGCCCATGAGCTCATCAAGGAGAAAGACTAAGTCAAGAGATGCC 1082  
 DB 752 --LeuSerHisProSerValGlnGlu--AlaSerSerAlaSerPheSerLeuSerSerPro 770  
 QY 1083 TCTCTGACCTTACAGATTCCTCTCTCCCTCCGACGCTTGATCTTG 1127  
 DB 771 ValMetThr-----SerSerSerProValSerSerThrLeu 782

## RESULT 14

US-09-925-297-668

Sequence 668: Application US/09925297

Patent No. US20020081659A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA105

CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: US/09/925,297

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 928

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 668

LENGTH: 566

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (178)

OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (357)

OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (518)

OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids

US-09-925-297-668

Alignment Scores:

Pred. No.: 0.756 Length: 566  
 Score: 111.00 Matches: 102  
 Percent Similarity: 34.68% Conservative: 52  
 Best Local Similarity: 22.97% Mismatches: 179

Query Match: 2.47% Indels: 111  
 DB: 10 Gaps: 19  
 US-09-806-276a-4 (1-2546) x US-09-925-297-668 (1-566)  
 QY 131 GCCCCCGGAGGCGCCGCGGCGGAGCTGGAGATCTTCACGCGGCGCGGTTCGTGGTT 190  
 DB 4 AlasergGluValValValValValValValValValValValValValValValVal 11e 20  
 QY 191 GTGGTCAGACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 250  
 DB 21 ValValLys-----AsnaProaspGlnSerValThrAla 32  
 QY 251 GGAAGTCTCTTTGACCCCATCTCATTTGCAAGC-----286  
 DB 33 AsnaAlaAlaThrAsnSerProSerGlyThrAlaAlaProProlValAsnValSerIlePro 52  
 QY 287 ACTTTCAGAGCCACTTGAGAAAATGATGACAGTCCATCAAAAAGATTTCAGAAC 346  
 DB 53 SerSerGluAlaLeuSerSer-----AspProSerTyrAsnLysGlu---LysHis 68  
 QY 347 ATATACCATCTGTGAAGAAGTGGCCCTTCTCCCGCTTGCAGAAATGACATTCATCAAT 406  
 DB 69 IleIleHis-----ProThrGlnLys-----75  
 QY 407 TCCAAATGCCAGCCAGACCCCAATTTACTGAAAAGCCCAAT-----ACAAGAAA 460  
 DB 76 SerLysAlaSerGlnLysSerSerSerSerSerSerSerSerSerSerSerSerSerSer 95  
 QY 461 GGAAG-----AAATTAACGTGAGGAGACATTCGTCTCTGATGATGATC 505  
 DB 96 LysLysGluLysSerThrSerLysTyrGluValLeuThrValGlnGluProProArgIle 115  
 QY 505 -----505  
 DB 116 GluAspAlaGluGluInPheProaspLeuAlaValAlaSerGluArgArgAspArgIleGlu 135  
 QY 506 -----AGTCCCGCGCTTGGAGACTTGGCCACACATCCACAT 544  
 DB 136 ThrProLysPheGlnSerLysGlnGlnProGlnAspAsnPhelLysAsnAsnVal-----153  
 QY 545 GGCAGAGAGGCGCAGCAGCATGCTTTCGAGATATTCCTTCTTCAGAGCAATCAGAG 604  
 DB 154 ---LysLysSerGlnLeuProValGlnLeuAsp-----LeuGlyLysMetLeuThr 169  
 QY 605 CTTTACCTCGAAGAACAGCAAGCAACCTCGGCGCAGTCC-----CTGGGCAAT 655  
 DB 170 AlaLeuGluLysLysGlnHisSer\*\*\*HisAlaLysGln--SerSerLysProValValVa 189  
 QY 189 lserValGlyAlaValProValLeuSerLysGlnLysAlaSerGlyGluArgGlyArgArg 209  
 QY 716 GTGCTCAAAAATGCCATCTCCCTCCGACCATGAGAGATCCCAACTCTCATGTGGCC 775  
 DB 209 gmetSerGlnMetLysThrProHisasnProLeuasp--SerSerAla-----224  
 QY 776 TTATTCACCAAGTACATTTAATCCAAACAGAGATCTCTCGGCGCAGCAAGATGCCCC 835  
 DB 225 -----ProLeuMetLysLysGlyLysGlnArgGlnIleProLysAlaLysLysPro 241  
 QY 836 AGCGTTAGTCGAGCCCGTCATAGAGAAAAGCTCAGAGAAAAGCAGTCTGTTGGAG 895  
 DB 242 ThrSerLeuLysLysIleIleLeuLysGlnArgGlnLysGlnArgGlnLysGlnGln 261  
 QY 896 AATGG-----ACAGTCCACAGGAGAGACACCTCGGGCGTCCAGCGTGTGGA 946  
 DB 262 AsnaIaValSerProAlaPheThrSerAspAspThrGlnAspLysLysSerGlyLysAsp 281  
 QY 947 TCTCAG---TCCAGCCAGCAGACAGACAGCAGCAGCAGCAGCAGCAGCAGCAGTACCC 1003  
 DB 282 AspGlnThrProGlnGlnAlaGlnLeuSerGlyProGlnLysMetaspGlnLeuIleSer 301

```

QY 1004 GACTGCCCGCAGGACATGTTTACCATCCACCCCATGCGATCATCAAGGAAAG 1063
      |||||
      :|||
      :|||
Db 302 ThrProSerValGluAspLysSerGluGluProProGlyThrGluLeuGlnArgAspThr 321
      |||||
      :|||
      :|||
QY 1064 ACTAGTCA-----GAGGAGTCCCTC 1084
      |||
      :|||
      :|||
Db 322 GluAlaSerHisLeuAlaProAsnHisThrThrPheProLysIleHisSerArgArgPhe 341
      |||
      :|||
      :|||
QY 1085 TCTGACCTTACAGGTTCCCTCTCTCTCTCTGCGATGAT---CTTGCGCCCTCACTTTG 1141
      |||
      :|||
      :|||
Db 342 ArgAspLysCysSerGlnMetLeuSerLysGluValAspAlaLys**ThrAspLeuLeu 361
      |||
      :|||
      :|||
QY 1142 GATAGCTGCTGAATGTAATGATTAATAATAGTACACAGATGCCAATTCTTCTG 1201
      |||
      :|||
      :|||
Db 362 LysGluLeuValArgPheGlnAspArgMetLysGlnLysasp-----ProVal 377
      |||
      :|||
      :|||
QY 1202 GGGTAAAGGTACAAACAACTAACCACAGTTGAAGACAGGCTTCCGAGCTGTAT 1261
      |||
      :|||
      :|||
Db 378 LysAlaLysThrLysAlaGluValLeuGlyLeuArgGluValLeuLysHisLeuLys 397
      |||
      :|||
      :|||
QY 1262 TTGCAG 1267
      |||
      :|||
      :|||
Db 398 LeuLys 399

```

## RESULT 15

```

US-09-864-761-36198
: Sequence 36198, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aeomica-x-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203

```

```

: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
: SEQ ID NO 36198
: LENGTH: 223
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC002041.1
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.4
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.8
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.9
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.2
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.7
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.2
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.3
: OTHER INFORMATION: EST HUMAN HIT: BE982238.1, EVALUE 2.00e-04
: OTHER INFORMATION: SWISSPROT HIT: P08775, EVALUE 6.00e-06
US-09-864-761-36198

```

## Alignment Scores:

```

Pred. No.: 0.67 Length: 223
Score: 110.00 Matches: 56
Percent Similarity: 35.75% Conservative: 23
Best Local Similarity: 25.34% Mismatches: 80
Query Match: 2.44% Indels: 62
Gaps: 11

```

US-09-806-276a-4 (1-2546) x US-09-864-761-36198 (1-223)

```

QY 522 ACTTTCGCCACACATCCATTTGGCAAGAGGCC-----ACCACGATCTTTG 572
      |||||
      :|||
      :|||
Db 4 ThrPheProArgSerProHisMetSerHisArgSerProHisThrSerProArgSerPro 23
      |||||
      :|||
      :|||
QY 573 GAGATATTCTCTTCTTCAAGGAACTACAGAGCTTTTACTCGAAACAGCAAGAACAC 632
      |||||
      :|||
      :|||
Db 24 HisThrPhePro-----ArgSerProHis 31
      |||||
      :|||
      :|||
QY 633 ACCTGGGCGACATTCCTGGGCGATGATGATTCCTCCGGGCAACAGCAGCTCGAGCTG 692
      |||||
      :|||
      :|||
Db 32 ThrPheProArgSerProHisThrSerProArgSerProHisThrSerProArgSerPro 51
      |||||
      :|||
      :|||
QY 693 TGTTCACAGAAAGCCCTCCCGG-----TGTCAAAATGCCATCTCCCTCCGA 743
      |||||
      :|||
      :|||
Db 52 HisThrPheProArgSerProHisThrPheProThrSerArgHisThrSerProArgSer 71
      |||||
      :|||
      :|||
QY 744 CCATTGGAGATCCACAGCTCTCATGTTGCCCTTATGT-----CACAG 788
      |||||
      :|||
      :|||
Db 72 ProHisThrPhePro-----ThrSerProHisThrSerAlaThrSerProHisMet 88
      |||||
      :|||
      :|||
QY 789 TGACATTAAATTCGAACAGAGACTCTTCGGGCGACCAAGCTGCCAGGCTTAGCTGGC 848
      |||||
      :|||
      :|||
Db 89 SerHisArgSerProHisThrSerProArgSerProHisThrSerPro-----104
      |||||
      :|||
      :|||
QY 849 AGCCGCTCATGAGGAAAGAGTCAGGAGAAAGCACTGTGTGGAGATGGACAGCTCC 908
      |||||
      :|||
      :|||
Db 105 -----ArgSer 106
      |||||
      :|||
      :|||
QY 909 ACCAGGAGACACCTCTGGGGCTCCAGCGTCTGCATCTCAGTCCAGCAGGAGAGAG 968
      |||||
      :|||
      :|||
Db 107 ProHisThrSerProArg-----SerProHisMetSerHisArgSerPro-----121
      |||||
      :|||
      :|||
QY 969 ACAGCCACTCTCCAGCTCTGTCCGAAAC-----AGTACCCGAGTGGCAGCGAGAG 1019
      |||||
      :|||
      :|||
Db 122 ---HisThrSerProArgSerProHisThrPheAlaThrSerProHisThrTyrrProArg 140
      |||||
      :|||
      :|||
QY 1020 ACATGTTTGACCATCCCA-----GCCCATGGCAGCTCATCAAGGAAAGACTAAGTACG 1073
      |||||
      :|||
      :|||
Db 141 SerProHisThrPheProHisThrSerProHisThrSerAlaThrSerProHisMetSerHis 160
      |||||
      :|||
      :|||

```

QY 1074 AGGAGTCCCTCTCTGACCTTACAGGTTCCCTCCCTCCCTGACGCTTGATCTTG6CCCT 1133  
 |||||  
 Db 161 ArgSerProHis-----ThrTyrProArgSerProHisThrSerProArgSerPro 177  
 |||||  
 QY 1134 CAC 1136  
 |||||  
 Db 178 His 178

Search completed: July 18, 2003, 17:44:58  
 Job time : 136.076 secs

472  
413





```

ID      08VC16      PRELIMINARY;      PRT;      238 AA.
OC      08VC16
DT      01-MAR-2002 (TREMBLrel. 20, Created)
DT      01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Hypothetical 26.2 kDa protein.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=COLON;
RA      Strausberg R.;
RL      Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC019760; AAH19760.1; -
DR      InterPro: IPR003589; Ig
DR      InterPro: IPR003597; Ig_c1
DR      InterPro: IPR003006; Ig_MHC.
DR      InterPro: IPR003596; Ig_v.
DR      Pfam: PF00047; Ig; 2.
DR      SMART: SM00409; Ig; 2.
DR      SMART: SM00407; IGH1.1.
DR      SMART: SM00406; IGV; 1.
DR      PROSITE: PS00290; Ig_MHC; UNKNOWN_1.
KW      Hypothetical protein.
SQ
SEQUENCE      238 AA; 26224 MW; 35FC08A730F541A4D C9064.

```

Alignment Scores:	
Pred. No.:	2,36e-59
Score:	737.00
Percent Similarity:	76.15%
Best Local Similarity:	57.32%
Query Match:	42.26%
DB:	11
	Gaps:
	Indels:
	6
	2

[illegible]

OY	579	TCGGGTACTCCAGGACAGTCTCACAGCAGCAGACAGAAGACAGCACTTACAGCTC	638
Dd	180	GlnAsnGlyValLeuAsnSerTrpThrAspAlaAspSerLysAspSerThrTySerMet	199
OY	639	AGCAGCACCCTGACGCCTGACGAAGCGACGACGAGAGAACACAAGTCTACCGTGAA	698
Dd	200	SerSerThrLeuThrLeuThrLysAspIleTyrGluArgHisAsnSerTyrThrCysGlu	219
OY	699	GTCACCATTCAGGGCCTGACCTGCCCGCTGCACAAGAGCTTCACAGGGGAGAGT	755
Dd	220	AlaThrHisLysThrSerThrSerProIleValLysSerPheAsnArgAsnGluCys	238
RESULT 4			
O8VCP0	ID	O8VCP0	PRELIMINARY; PRT; 234 AA.
AD	O8VCP0;		
DT	01-MAR-2002 (TREMBL)	20	Created
DE	01-MAR-2002 (TREMBL)	20	Last sequence update
DE	01-JUN-2002 (TREMBL)	21	Last annotation update
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
NCBI_OXID	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	TISSUE=COLON:		
RL	Submitted by R. Strussberg R.		
DR	EMBL: BC019474; AAI19474.1; -		
DR	InterPro: IPR003599; Ig.		
DR	InterPro: IPR003597; Ig_C1.		
DR	InterPro: IPR003006; Ig_MHC.		
DR	InterPro: IPR003596; Ig_V.		
DR	Pfam: PF00047; Ig_2.		
DR	SMART: SMO0409; IG; 2.		
DR	SMART: SMO0407; IGc1; 1.		
DR	SMART: SMO0406; IgV; 1.		
KW	PROSITE; PS00290; IG_MHC; UNKNOWN_1.		
Hypotheical protein.			
SQ	SEQUENCE 234 AA; 25702 MW; 102551C58AC2FA9F CRC64;		
Alignment Scores:			
Pred. No.:	6,78e-59	Length:	234
Score:	732.00	Matches:	138
Best Similarity:	73.93%	Conservative:	35
Best Local Similarity:	58.97%	Mismatches:	61
Query Match:	41.97%	Indels:	0
Dd:	11	Gaps:	0
US-09-806-276A-3 (1-962) x O8VCP0 (1-234)			
OY	54	ATGGAAGCCCCAGCTCAGCTTCTCTCTCTGCTGCTCACTCTGGCTCCAGATACACCGGA	113
Dd	1	MetserValProThngInValLeuGlyLeuLeuLeuLeuLeuCysLeuThngIAlaArgCys	20
OY	114	GAATTGTGTGACACAGTCTCCAGCACCCTGCTTTGCTCTCCAGGGAAAGACCCACC	173
Dd	21	AspiIleGlnLeuThngInserProAlaserLeuSerAlaserValGlyGIuThrValTr	40
OY	174	CCTCCTCAGGGCCATCAGAGTGTTAGCAGCTACTTACGCTGTAACCAAGAGAACT	233
Dd	41	IleMrcySArGalaSerGluAsnIleTyrSerTyrLeuAlaIlePrtyrGlnGlnGln	60
OY	234	GGCAGGCTCCACAGGCTCCTCATCTATGATGATCATCAAGAGGGCCACTGGCATCCACC	293
Dd	61	GlyLysSerProGlnLeuLeuValTyrSnaIalylThrLeuAlaSpGlyValProser	80
OY	294	AGGTTCAAGTGGAGTGGCTGGGACACACTTCACTCACTCAACAGCAGACTGGAGCC	353
Dd	81	ArgPheSerIleYserIArgSerIylThrIlnPheSerLeuLysIleAsnSerLeuGlnPro	100





```

QY 114 GAAATGTGTGACAGCTGCTCCAGCCCTGTCTTGTCTCCAGGGGAAAGCCACC 173
Db 21 AsplegmetrhrlnthrhSerSerleuSerlaSerleuGlyAsparValThr 40
QY 174 CTCTCTGCGAGCGCCAGTCAAGTGTAGACCTTACTAGCTGTGTACCAAGAACT 233
Db 41 IleserCysArglaSerlInspIleserlnsPryleuasnPrtyrGlnGlyPro 60
QY 234 GGGCAGGCTCCAGGCTCTCATATGATGATCCACAGAGGCGCATCCGACCC 293
Db 61 AspGlyThrVallyleuSerleuSerleuSerleuSerleuSerleuSerleu 80
QY 294 AGGTTACAGGCGAGTGGTCTGCGACAGCTTCACTCCACAGACAGCTGAGCCC 353
Db 81 ArgPheSerGlySerGlySerGlyThrAspPrySerleuThrIleserlnsGln 100
QY 354 GAAGATGTGCGACTTATTACTGACCAATATTATTACTGCTGCTGCTGCTGCTG 413
Db 101 GluAspIleAlaThrTyrrPheCysGlnGlnGlyAsnThrProPheThrPheGlySer 120
QY 414 GGGACAGGCTGCGAGATCAACAGAACTGTGCTGACCAATGCTGCTGCTGCTG 473
Db 121 GlyThrLysLeuGlnValLysArgAlaAspAlaAlaProThrValSerleuPhePro 140
QY 474 TCTGATGAGCTGGAATCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 533
Db 141 SerSerGlnGlnLeuThrSerGlyGlyAlaSerValValCysPheLeuAsnAsnPhet 160
QY 534 CCCAGAGCGCCAAAGTACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 593
Db 161 ProlAspIleAsnValLysTrpLysIleAspGlySerGlnValGlnAsnGlyVal 180
QY 594 GAGAGTGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 653
Db 181 AsnSerTrpThrAspIleAsnSerLysAspSerThrTyrrSerMetSerThrLeu 200
QY 654 CTGAGCAAGCAGACTACAGCAAAAGTCTAGCGCTGCGAAGTCAACCCATCGAG 713
Db 201 LeuThrLysAspGlyIleGlnValGlnAsnSerLysGlnValAlaThrHisLys 220
QY 714 CTGAGCTCCCGCTCAAGAGAGCTTCAACAGGAGAGAGTGT 755
Db 221 SerThrSerProIleValLysSerPheAsnArgAsnGlnLys 234

```

RESULT 7  
Q8VC55  
AC 08VC55; PRELIMINARY; PRT; 239 AA.  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN NCBI\_TaxID=10090;  
RP SEQUENCE FROM N.A.  
RC TISSUE=COLON;  
RA Strausberg R.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC021781; AAK21781.1;  
DR InterPro: IPR003599; Iq.  
DR InterPro: IPR003597; Iq\_c1.  
DR InterPro: IPR003006; Iq\_MHC.  
DR InterPro: IPR003596; Iq\_v.  
DR Pfam: PF00047; Iq\_2.  
DR SMART: SM00409; Iq\_1.  
DR SMART: SM00407; Iqcl\_1.  
DR SMART: SM00406; Iq\_1.  
DR PROSITE: PS00290; Iq\_MHC; UNKNOWN\_1.  
KW Hypothetical protein.

```

SQ SEQUENCE 239 AA; 26303 MW; C16119CACA25C337 CRC64;
Alignment Scores:
Pred. No.: 9.71e-58
Score: 719.50 Length: 239
Best Local Similarity: 74.90% Matches: 136
Best Local Similarity: 56.90% Conservative: 43
Query Match: 41.26% Mismatches: 55
DB: Gaps: 5
US-09-806-276a-3 (1-962) x Q8VC55 (1-239)
QY 54 ATGAGACCCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 113
Db 1 MetLysLeuProValLeuLeuValValLeuLeuLeuPheThrSerProAlaSerSer 20
QY 114 GAAATGTGTGACAGCTGCTCCAGCCCTGTCTTGTCTCCAGGGGAAAGCCACC 173
Db 21 AspValValLeuThrGlnThrProLeuSerleuProValAsnIleGlyAspAlaSer 40
QY 174 CTCTCTGCGAGCGCCAGTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 218
Db 41 IleserCysLysSerThrLysSerleuLeuAsnSerLysPheThrValSerAsn 60
QY 219 TACCAACAGAAAGCTGCGAGGCTCCAGGCTGCTGCTGCTGCTGCTGCTGCTG 278
Db 61 TyrrLeuGlnLysProGlyGlnSerProGlnLeuLeuLeuLeuValSerAsnArg 80
QY 279 ACTGCAATCCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 338
Db 81 SerGlyValProAlaPargPheSerGlySerGlySerGlyThrAspPheThrLeu 100
QY 339 AGCAGACTGAGAGCCCGAGAGTGGCACTTATTACTGCTGCTGCTGCTGCTGCTG 398
Db 101 SerArgValGlnAlaGlnAspLeuGlyValTyrrTyrrCysPheGlnSerAsnHis 120
QY 399 TACACTTTTGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 458
Db 121 TyrrThrPheGlyIleGlyThrLysLeuGlnIleLysArgAlaAspAlaAlaPro 140
QY 459 TTCATCTTCCCGCATCTGATGATGATGATGATGATGATGATGATGATGATGAT 518
Db 141 SerIlePheProPheSerSerGlnGlnLeuThrSerGlyAlaSerValValCys 160
QY 519 CTGAATTAATCTTATCCAGAGAGGCAAGTACAGTGAAGTGAAGTGAAGTGAAG 578
Db 161 LeuAsnAsnPhetLysProlLysAspIleAsnValLysTrpLysIleAspGlySer 180
QY 579 TCGGGTAATCCAGAGAGTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638
Db 181 GlnAsnGlyValLeuAsnSerTrpThrAspIleAsnSerLysAspSerThrLys 200
QY 639 AGCAGACCCCTGAGCTGAGCAAGAGTACAGAGAGAGAGAGAGAGAGAGAGAG 698
Db 201 SerSerThrLeuThrLeuThrLysAspGlyIleGlnValGlnAsnSerLysGln 220
QY 699 GTACACCATCAGGCGCTGAGGCTGCGCTGCAAGAGCTTCAACAGGAGAGAGT 755
Db 221 AlaThrHisLysThrSerThrSerProIleValLysSerPheAsnArgAsnGln 239

```

RESULT 8  
Q91MS9  
AC 091MS9; PRELIMINARY; PRT; 233 AA.  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN NCBI\_TaxID=10090;  
[1]

us-09-806-276a-3.rsp

RP	SEQUENCE FROM N.A.	
RC	TISSUE=COLON;	
RA	Strasberg R.;	
RL	Submitted (SEP-2001) to the EMBL/GenBank/DBDj databases.	
RL	EMBL: BC013496; AAH13496.1.	
DR	InterPro: IPR003006; I9_MHC.	
DR	Pfam: PF00047; I9; 2.	
DR	PROSITE: PS00290; I9_MHC; UNKNOWN_1.	
KM	Hypothetical protein.	
FT	NON_TER	1
SQ	SEQUENCE	233 AA; 25781 MW; BIC184DA149A16BB CRC64;
Alignment Scores:		
Align. No.:	3, 21e-56	Length: 233
Score:	703.00	Matches: 132
Percent Similarity:	74.55%	Conservative: 35
Best Local Similarity:	58.93%	Mismatches: 57
Query Match:	40.31%	Indels: 0
DB:	11	Gaps: 0

```

US-09-806-276A-3 (1-962) x Q91WS9 (1-233)

QY      84 CTGCTACCTCGTGGCTCCCGAGATATCCACCGCGAGAAATTGTGTACACAGTCTCCAGCGCAC 143
          |||||
          .....
Db      10 LeuLeuLeuCySpheInglnglySerArgCysaspLleGlnMetThrGlnThrThrSerSer 29
          |||||
          .....
QY      144 CTGCTTTGTCTCCAGCGGAAAGACCCACCTCTCTCGCAGGGCCATCGAGCTGTTCAG 203
          |||||
          .....
Db      30 LeuSerAlaSerLeuGlnGlyAspArgValThrIleSerYsserGlySerGlnGlyIleAla 49
          |||||
          .....
QY      204 AGCTACTTGACCTGGTACCAACAGAAACCTGGCCAGCTCCCGAGGCTCCATCTCATGAT 263
          |||||
          .....
Db      50 AsnTyrLeuAsnTrpTyrGlnGlnLysProAspGlyThrValLysLeuIleTyrTyr 69
          |||||
          .....
QY      264 GCATCCACAGGGCCACTGGCATCCACCCAGGTTCAGTGCAGTGGCTGGAGACAGC 323
          |||||
          .....
Db      70 ThrSerSerLeuHisSerIlyValProSerArgPheSerGlySerGlyThrAsp 89
          |||||
          .....
QY      324 TTCACGTCCACATCAGCAGACTGGAGCCCGAAGATGTGGCACTTATTACGTGCAGCAA 383
          |||||
          .....
Db      90 TyrSerLeuThrIleSerAsnLeuGlnProGluAspLleAlaThrTyrTyrCysGlnGln 109
          |||||
          .....
QY      384 TATTTTACTACTCGCGTAACACTTTTGGCCAGGGGACGAGCTGGAGATCAACGAAGTGG 443
          |||||
          .....
Db      110 TyrIrrGlyLeuProIrrPrrPheGlyGlyGlyThrLysLeuGlnIleLysArgAlaAsp 129
          |||||
          .....
QY      444 GCTCCACCATCTGTCTTCATCTTCCGCCCATCTGATGAGAGAGTGAATCTGAACTGGCC 503
          |||||
          .....
Db      130 AlaAlaProIrrValSerIlePheProProSerSerIleuLneuthrSerGlyGlyAla 149
          |||||
          .....
QY      504 TCTGTGTGGCGCTGCGAATAACTTATATCCAGAGCGCAAAAGTACAGTGGAAAGGTG 563
          |||||
          .....
Db      150 SerValValLysPheLeuAsnAsnPhetyrProLysaspLleAsnValLysTrpLysIle 169
          |||||
          .....
QY      564 GATTAACGCGCTCCAAATCGGTAACTCCAGAGAGGTGTCCACAGCGAGACACAGAGAC 623
          |||||
          .....
Db      170 AspGlySerGluArgGlnAsnGlyValLeuAsnSerTrpPrrAspGlnAspSerLysAsp 189
          |||||
          .....
QY      624 AGCAGCTTAAGCGCTACAGGACCGCTGACGCTGAGCAAAAGCAGACTACGAGAAACAAA 683
          |||||
          .....
Db      190 SerThrTyrSerMetSerSerThrIleuthrIleuthrLysAspGlyGlyAlaArgHisAsn 209
          |||||
          .....
QY      684 GTTACAGCTGGGAAGTACCCATCAGAGGCGCTGAGAGCTGGCCGCGTACCAAAAGAGTTAC 743
          |||||
          .....
Db      210 SerTyrThrCysGlnAlaThrHisLysThrSerThrSerProIleValLysSerPheAsn 222
          |||||
          .....
QY      744 AGGGAGAGAGTCT 755
          |||||
          .....
Db      230 ArgAsnGluCys 233
          |||||
          .....

RESULT 9
Q9RLAS PRELIMINARY; PRT; 214 AA.
ID Q9RLAS

```

Alignment Scores:	
pred. No.:	7.32e-56
Score:	699.00
Percent Similarity:	76.64%
Best Local Similarity:	59.35%
Query Match:	40.08%
DB:	11
Length:	214
Matches:	127
Conservative:	37
Mismatches:	50
Indels:	0
Gaps:	0

[illegible]

[illegible]



## RESULT 14

### Alignment Scores:

US-09-806-276A-3 (1-962) x Q8WTU6 (1-237)

US-09-806-2/6A-3 (1-962) x Q8WTU6 (1-237)

81 CTCCTGCTCTGGCTCCAGATACCAACCGA-----GAAATTGTGTTGACACAC 131

[illegible]

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

1997

44 In the case of the cysteine protease, the amino acid sequence of the protein was determined by mass spectrometry and compared with the known sequence of the cysteine protease from the same organism.

192 CAGAGT-----GTTAGCAGCTACTTAGCCTGGTACCAACAGAAACCTGGCCAGGCT 242

45 SerSerAsnIleGlyAlaGlyTyrAspValHisTrpTyrGlnGlnLeuProGlyThrAla 64

243 CCCAGGCTCCTCATCTATGATGCATCCAAACAGGGCAATTGGGTAATCCCCCCCCTTTACCC

```
65 PROTOCOL OUT PUT COMMANDS
```

65 ProLysLeuLeuIleTyrglyAsnSerAsnArgProSerGlyValProAspArgPheSer 84

QY	303	GGGAGTGGGCTGGGACAGCTTCCTTCACATCAGCAGACAGCCGAGAGTGTG	367
Db	85	GLYserIysserGIythrSerIasIerLeuAlaIlethrGIyLeuGlnAlaGluAspLeu	104
QY	363	GCACCTATTACTGTGCAGCAATAT-----TTTACTACCTCCACACTTTGGCCAG	413
Db	105	AlaAspIyTrIyCysInSerIyTrAspIyTrSerLeuSerIasIerGIyValAlaPheGIy	124
QY	414	GGGAGCCAGGCTGGAGATC--AAAGCACTGTGGCTGCACATCTGTCTTATCTCCCG	470
Db	125	GIyThrIyLysLeuThrValLeuGIyGlnProIySAIAlaIaProSerValThIleuPhePro	144
QY	471	CCATCTGATGAGGACGTTGAATCTGGAACTGCCCTGTGTGTGGCTGCAGTAATACCTC	530
Db	145	ProSerSerGIuGIuLeuGlnAlaAsnIySAIatThrLeuValCysLeuIleSerAspPhe	164
QY	531	TATCCAGAGAGCCCAAGTACAGTGGAGAGGTGATTAAGCC--CTCCAAVCGGGTAAAC	587
Db	165	TyPProGIyAlaValAlaThrValAlaIatTrpIySAIAspSerProValIyAlaGIy---	183
QY	588	TCCAGAGAGGTGTCACAGAGCAGGACGACCAAGCACCACCTACAGCTTAGACACC	647
Db	184	---ValGIuThrThrThrProSerIyGlnSerAsnAsnIyTrYAlaAlaIleSerSerTy	202
QY	648	CTGAGCCTAGACCAAGACAGACCTACAGAAACAACAAGCTCAGGCTCGGAATACACC	707
Db	203	LeuSerIeuthrProGIuGIuTrpIySerSerIAsIySerIyTrSerCysGIuValThHis	222
QY	708	CAAGGCGCTAGCTCGCCCTGCACAAAGAGCTTCAACAGGGAGAGTGT	755
Db	223	GIuGIyL-----SerThrValGIuIySThrValAlaIaProThrGIuIyS	236

## RESULT 15

ID	09U166;	PRELIMINARY;	PRT;	109 AA.
AC	09U166;			
DT	01-MAY-2000 (TREMBLrel, 13, Created)			
DT	01-MAY-2000 (TREMBLrel, 13, last sequence update)			
DT	01-DEC-2001 (TREMBLrel, 19, last annotation update)			
DE	Myosin-reactive immunoglobulin kappa chain variable region (Fragment).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
ON	[1]			

NP SEQUENCE FROM N.A.  
RX MEDLINE=98277139; PubMed=9614934;  
RA Wu X., Liu B., Van der Werf P.L., Kalis N.N., Berney S.M.,  
RA Young D.C.;  
RT "Myosin-reactive autoantibodies in rheumatic carditis and normal  
RT fetus";  
RL Clin. Immunol. Immunopathol. 87:184-192(1998).  
EMBL: AF035028.2.DD56266

DR HSSP; P80362; 1WTL.

DR InterPro; IPR003596; Ig\_v.  
DR Pfam: PF00047.1

DR SMART; SM00406; IGV; 1.

ET	1	109
NON-TER	1	109

SEQUENCE 109 AA; 11928 MW; 243325F72C7DAC83 CRC64;

Alignment Scores:

Score:	453.50	Matches:	90
Length:	109		

Percent Similarity:	89.91%	Conservative:	8
Best Total Similarity:	90.57%	Conservative:	8

```
Query Match: 26.00% Indels: 1
```

1. *Chlorophyll a* (Chl *a*)  
 2. *Chlorophyll b* (Chl *b*)  
 3. *Chlorophyll c* (Chl *c*)  
 4. *Chlorophyll d* (Chl *d*)  
 5. *Chlorophyll e* (Chl *e*)  
 6. *Chlorophyll f* (Chl *f*)  
 7. *Chlorophyll g* (Chl *g*)  
 8. *Chlorophyll h* (Chl *h*)  
 9. *Chlorophyll i* (Chl *i*)  
 10. *Chlorophyll j* (Chl *j*)  
 11. *Chlorophyll k* (Chl *k*)  
 12. *Chlorophyll l* (Chl *l*)  
 13. *Chlorophyll m* (Chl *m*)  
 14. *Chlorophyll n* (Chl *n*)  
 15. *Chlorophyll o* (Chl *o*)  
 16. *Chlorophyll p* (Chl *p*)  
 17. *Chlorophyll q* (Chl *q*)  
 18. *Chlorophyll r* (Chl *r*)  
 19. *Chlorophyll s* (Chl *s*)  
 20. *Chlorophyll t* (Chl *t*)  
 21. *Chlorophyll u* (Chl *u*)  
 22. *Chlorophyll v* (Chl *v*)  
 23. *Chlorophyll w* (Chl *w*)  
 24. *Chlorophyll x* (Chl *x*)  
 25. *Chlorophyll y* (Chl *y*)  
 26. *Chlorophyll z* (Chl *z*)  
 27. *Chlorophyll aa* (Chl *aa*)  
 28. *Chlorophyll ab* (Chl *ab*)  
 29. *Chlorophyll ac* (Chl *ac*)  
 30. *Chlorophyll ad* (Chl *ad*)  
 31. *Chlorophyll ae* (Chl *ae*)  
 32. *Chlorophyll af* (Chl *af*)  
 33. *Chlorophyll ag* (Chl *ag*)  
 34. *Chlorophyll ah* (Chl *ah*)  
 35. *Chlorophyll ai* (Chl *ai*)  
 36. *Chlorophyll aj* (Chl *aj*)  
 37. *Chlorophyll ak* (Chl *ak*)  
 38. *Chlorophyll al* (Chl *al*)  
 39. *Chlorophyll am* (Chl *am*)  
 40. *Chlorophyll an* (Chl *an*)  
 41. *Chlorophyll ao* (Chl *ao*)  
 42. *Chlorophyll ap* (Chl *ap*)  
 43. *Chlorophyll aq* (Chl *aq*)  
 44. *Chlorophyll ar* (Chl *ar*)  
 45. *Chlorophyll as* (Chl *as*)  
 46. *Chlorophyll at* (Chl *at*)  
 47. *Chlorophyll au* (Chl *au*)  
 48. *Chlorophyll av* (Chl *av*)  
 49. *Chlorophyll aw* (Chl *aw*)  
 50. *Chlorophyll ax* (Chl *ax*)  
 51. *Chlorophyll ay* (Chl *ay*)  
 52. *Chlorophyll az* (Chl *az*)  
 53. *Chlorophyll aza* (Chl *aza*)  
 54. *Chlorophyll abz* (Chl *abz*)  
 55. *Chlorophyll acz* (Chl *acz*)  
 56. *Chlorophyll adz* (Chl *adz*)  
 57. *Chlorophyll aez* (Chl *aez*)  
 58. *Chlorophyll afz* (Chl *afz*)  
 59. *Chlorophyll agz* (Chl *agz*)  
 60. *Chlorophyll ahz* (Chl *ahz*)  
 61. *Chlorophyll aiz* (Chl *aiz*)  
 62. *Chlorophyll ajz* (Chl *ajz*)  
 63. *Chlorophyll akz* (Chl *akz*)  
 64. *Chlorophyll alz* (Chl *alz*)  
 65. *Chlorophyll amz* (Chl *amz*)  
 66. *Chlorophyll anz* (Chl *anz*)  
 67. *Chlorophyll aoz* (Chl *aoz*)  
 68. *Chlorophyll apz* (Chl *apz*)  
 69. *Chlorophyll aqz* (Chl *aqz*)  
 70. *Chlorophyll arz* (Chl *arz*)  
 71. *Chlorophyll asz* (Chl *asz*)  
 72. *Chlorophyll atz* (Chl *atz*)  
 73. *Chlorophyll auz* (Chl *auz*)  
 74. *Chlorophyll avz* (Chl *avz*)  
 75. *Chlorophyll awz* (Chl *awz*)  
 76. *Chlorophyll axz* (Chl *axz*)  
 77. *Chlorophyll ayz* (Chl *ayz*)  
 78. *Chlorophyll ayz* (Chl *ayz*)  
 79. *Chlorophyll azz* (Chl *azz*)  
 80. *Chlorophyll azaa* (Chl *aza*)  
 81. *Chlorophyll abz* (Chl *abz*)  
 82. *Chlorophyll acz* (Chl *acz*)  
 83. *Chlorophyll adz* (Chl *adz*)  
 84. *Chlorophyll aez* (Chl *aez*)  
 85. *Chlorophyll afz* (Chl *afz*)  
 86. *Chlorophyll agz* (Chl *agz*)  
 87. *Chlorophyll ahz* (Chl *ahz*)  
 88. *Chlorophyll aiz* (Chl *aiz*)  
 89. *Chlorophyll ajz* (Chl *ajz*)  
 90. *Chlorophyll akz* (Chl *akz*)  
 91. *Chlorophyll alz* (Chl *alz*)  
 92. *Chlorophyll amz* (Chl *amz*)  
 93. *Chlorophyll anz* (Chl *anz*)  
 94. *Chlorophyll aoz* (Chl *aoz*)  
 95. *Chlorophyll apz* (Chl *apz*)  
 96. *Chlorophyll aqz* (Chl *aqz*)  
 97. *Chlorophyll arz* (Chl *arz*)  
 98. *Chlorophyll asz* (Chl *asz*)  
 99. *Chlorophyll atz* (Chl *atz*)  
 100. *Chlorophyll auz* (Chl *auz*)  
 101. *Chlorophyll avz* (Chl *avz*)  
 102. *Chlorophyll awz* (Chl *awz*)  
 103. *Chlorophyll axz* (Chl *axz*)  
 104. *Chlorophyll ayz* (Chl *ayz*)  
 105. *Chlorophyll ayz* (Chl *ayz*)  
 106. *Chlorophyll azz* (Chl *azz*)  
 107. *Chlorophyll azaa* (Chl *aza*)  
 108. *Chlorophyll abz* (Chl *abz*)  
 109. *Chlorophyll acz* (Chl *acz*)  
 110. *Chlorophyll adz* (Chl *adz*)  
 111. *Chlorophyll aez* (Chl *aez*)  
 112. *Chlorophyll afz* (Chl *afz*)  
 113. *Chlorophyll agz* (Chl *agz*)  
 114. *Chlorophyll ahz* (Chl *ahz*)  
 115. *Chlorophyll aiz* (Chl *aiz*)  
 116. *Chlorophyll ajz* (Chl *ajz*)  
 117. *Chlorophyll akz* (Chl *akz*)  
 118. *Chlorophyll alz* (Chl *alz*)  
 119. *Chlorophyll amz* (Chl *amz*)  
 120. *Chlorophyll anz* (Chl *anz*)  
 121. *Chlorophyll aoz* (Chl *aoz*)  
 122. *Chlorophyll apz* (Chl *apz*)  
 123. *Chlorophyll aqz* (Chl *aqz*)  
 124. *Chlorophyll arz* (Chl *arz*)  
 125. *Chlorophyll asz* (Chl *asz*)  
 126. *Chlorophyll atz* (Chl *atz*)  
 127. *Chlorophyll auz* (Chl *auz*)  
 128. *Chlorophyll avz* (Chl *avz*)  
 129. *Chlorophyll awz* (Chl *awz*)  
 130. *Chlorophyll axz* (Chl *axz*)  
 131. *Chlorophyll ayz* (Chl *ayz*)  
 132. *Chlorophyll ayz* (Chl *ayz*)  
 133.

US-09-806-276A-3 (1-962) x 091186 (1-109)

```
QY      114  GAAATTGTGTGACAGCTCTCCAGCCACCTGTCTTTGTCTCCAGGGGAAAGCCACC 173
      |||||||
Db      1  GIUIIeValIeuThrInSerProGIYThrIeuSerIeuPheProGIYGIuArgAlaThr 20
      |||||||
QY      174  CTCTCTGCAGGGCCAGTCAGAGTGT---AGCAGCTACTTAAGCTGTACCAACAGAAA 230
      |||||||
Db      21  LeuSerCysArgIaSerGIInSerValSerSerSeryrIeuAlaTrpTYrGIInGIuys 40
      |||||||
QY      231  CTTGGCCAGGCTCCAGGCTCCATCTATGATGCAATCCACAGGGCCACTGGCATGCCA 290
      |||||||
Db      41  ProGIYGIuAlaProArgIeuIeuIeYrGIYThrSerSerArgAlaThrGIYIlePro 60
      |||||||
QY      291  CCCAGGTTCAAGTCAGAGTGGGTCTGGACAGACTTCACCTCACCACATCAGCAGACTGGAG 350
      |||||||
Db      61  AspArgPheSerGIYserGIYserGIuThrAspPheThrIeuThrIleSerArgIeuGIu 80
      |||||||
QY      351  CCCGAGAGATGTGCACCTTTATTTACTGTACAGCATATTTTACTACTCCGTACACTTTGGC 410
      |||||||
Db      81  ProGIuAspPheAlaValTYrTYrCysGIInGIuTYrGIYserSerIlePheThrPheGIY 100
      |||||||
QY      411  CAGGGACAGAGCTGGAGATCAACGA 437
      |||||||
Db      101  ProGIYThrIYsValAspIleIYsArg 109
```

Search completed: July 18, 2003, 17:34:47  
Job time : 74.1462 secs

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: July 18, 2003, 16:45:27 ; Search time 13.0259 seconds

6126.272 Million cell updates/sec

```

Title:      US-09-806-276A-3
Perfect score: 1744
Sequence:   1 tcgagccgattcgctcgag.....gaatctttgciaaaaaaaaa 962

```

blusun02	Xgapop	10.0	Xgapext	0.5
	Ygapop	10.0	Ygapext	0.5
	Fgapop	6.0	Fgapext	7.0
	Delop	6.0	Delext	7.0

----- 20907 444/0326 residues

Total number of hits satisfying chosen parameters: 225784

Post-processing: Minimum Match 0%

Listing first 45 summaries

```

Command line parameters:
-MODEL=framed_r2p.model -DEV=xlp
-O=cgml_1/USPTO.spool/US09806276/runat.18072003.160924.12742/app_query.fasta.1.38546
-DB=SwissProt_40 -QMT=fasta -SUFFIX=bsp -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=lsom62 -TRANS=humad40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORF=gen -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09806276.ecgn.1.1.79.gruntat.18072003.160924.12742 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGADEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DEPOP=6 -DELEXT=7

```

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
1	589.5	33.8	129	1	KV3L_HUMAN	P18135	homo sapien	
2	583.5	33.5	129	1	KV3M_HUMAN	P18133	homo sapien	
3	573	32.9	128	1	KV3K_HUMAN	P08311	homo sapien	
4	559.5	32.1	129	1	KV3L_HUMAN	P04207	homo sapien	
5	548	31.4	106	1	KAC_HUMAN	P01834	homo sapien	
6	545	31.2	115	1	KV3T_HUMAN	P04433	homo sapien	
7	490.5	28.1	116	1	KV3J_HUMAN	P04434	homo sapien	
8	479.5	27.5	109	1	KV3E_HUMAN	P01620	homo sapien	
9	477.5	27.4	109	1	KV3D_HUMAN	P01622	homo sapien	
10	470.5	27.0	109	1	KV3E_HUMAN	P01623	homo sapien	
11	456.5	26.2	108	1	KV3A_HUMAN	P01619	homo sapien	
12	453.5	26.0	109	1	KV3G_HUMAN	P04206	homo sapien	
13	444	25.0	129	1	KV3L_HUMAN	P04431	homo sapien	
14	442	25.3	134	1	KV4C_HUMAN	P06314	homo sapien	
15	439.5	25.2	109	1	KV3F_HUMAN	P01634	homo sapien	
16	427	24.5	131	1	KV3I_MOUSE	P01651	mus musculu	
17	424.5	24.3	100	1	KV3C_MOUSE	P01621	homo sapien	
18	423	24.3	132	1	KV3F_MOUSE	P01658	mus masculu	

19	422.5	24.2	133	1	KV4B_HUMAN	P06313	homo sapien
20	422	24.2	114	1	KV4A_HUMAN	P01625	homo sapien
21	419	24.0	117	1	KV1J_HUMAN	P01602	homo sapien
22	416.5	23.9	133	1	KV2E_HUMAN	P06310	homo sapien
23	410	23.5	108	1	KV1H_HUMAN	P01600	homo sapien
24	401	23.0	149	1	KV5A_MOUSE	P01653	mus musculus
25	400	22.9	108	1	KV1M_HUMAN	P01605	homo sapien
26	399	22.9	108	1	KV5E_MOUSE	P01637	mus musculus
27	399	22.9	128	1	KV5E_MOUSE	P04452	homo sapien
28	397	22.8	129	1	KV1X_HUMAN	P01544	homo sapien
29	396	22.8	108	1	KV1B_HUMAN	P04430	homo sapien
30	395	22.6	108	1	KV1A_HUMAN	P01593	homo sapien
31	395	22.6	108	1	KV1A_HUMAN	P01607	homo sapien
32	393	22.6	108	1	KV1O_HUMAN	P01607	homo sapien
33	392	22.5	121	1	KV4O_HUMAN	P06312	homo sapien
34	389	22.5	136	1	KV5B_MOUSE	P01634	mus musculus
35	388	22.3	108	1	KV1L_HUMAN	P01604	homo sapien
36	388	22.2	108	1	KV1E_HUMAN	P01597	homo sapien
37	388	22.2	108	1	KV1K_HUMAN	P01597	homo sapien
38	388	22.2	108	1	KV1N_HUMAN	P01606	homo sapien
39	385	22.2	108	1	KV1Y_HUMAN	P080362	homo sapien
40	383	22.1	108	1	KV1C_HUMAN	P01655	homo sapien
41	382.5	22.0	111	1	KV3H_MOUSE	P01660	mus musculus
42	381.5	21.9	117	1	KV2E_HUMAN	P06309	homo sapien
43	381.5	21.9	107	1	KV2D_HUMAN	P01596	homo sapien
44	381	21.8	113	1	KV2O_HUMAN	P01617	homo sapien
45	381	21.8	108	1	KV1O_HUMAN	P01605	homo sapien
			111	1	KV3B_MOUSE	P01665	mus musculus

## ALIGNMENTS

ID	KEY3L_HUMAN	STANDARD:	PRT:	129 AA.
AC	KEY3L_HUMAN			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Ig kappa chain V-II region HAM precursor.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=88171307; PubMed=3127527;			
RA	Kipps T.J., Tomhave E., Chen P.P., Carson D.A.;			
RT	"Autoantibody-associated kappa light chain variable region gene			
RT	expressed in chronic lymphocytic leukemia with little or no somatic			
RT	mutation. Implications for etiology and immunotherapy.";			
RL	J. Exp. Med. 167:840-852(1988).			
CC	-1- DISEASE: THE PROTEIN IS ONE OF THE SURFACE IMMUNOGLOBULIN M			
CC	AUTOANTIBODIES EXPRESSED IN PATIENTS WITH CHRONIC LYMPHOCYTIC			
CC	LEUKEMIA.			
DR	PIR; P00022; K3HCHA.			
DR	HSSP; P80362; 1WFL.			
DR	InterPro: IPR003006; Ig_MHC.			
DR	InterPro: IPR003596; 1g_V.			
DR	Pfam: PF00047; 1g; 1.			
DR	SMART; SM00406; IGV; 1.			
KM	Immunoglobulin V region; Signal.			
FT	SIGNAL	1		
FT	CHAIN	21	129	
FT	DOMAIN	21	43	
FT	DOMAIN	44	55	
FT	DOMAIN	56	70	
FT	DOMAIN	71	77	
FT	DOMAIN	78	109	
FT	DOMAIN	110	118	
FT	DOMAIN	119	129	
FT	DISULFID	43	109	
FT	NON_TER	129	129	
SEQUENCE	129 AA; 14073 MW; D3C5529227774D0 CRC64;			

44	55	COMPLEMENTARITY-DETERMINING-1.
44	55	DOMAIN
44	55	FT

CC use by non-profit institutions and for commercial  
CC modified and this statement is not removed. Usage by and for commercial  
CC modified and this statement is not removed. Usage by and for commercial

CC use by non-profit institutions as long as the original work is properly attributed. Usage by and for commercial entities and for other types of copying (e.g., mass reprints, advertising, promotional purposes, etc.) is not permitted without the express written permission of the publisher. Usage of the work in forms not intended by the publisher, such as photocopying, scanning, or otherwise, is not permitted without the express written permission of the publisher. For more information, contact the publisher or the Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA. CCC's website is at <http://www.copyright.com>.



NCBI\_TaxID=9606; Homo. catarrhini; Homnidae; Homo.

101 GluAsppheAlaValITTryCysGlnGlnIntYAsnAsntTrProTrProTrPTrTrPhesely 120  
411 CAGGGACACAGCGCTGAGATCAACGA 437

```

DB      121 GINGLYTHRARGVALGILUILEYSARG 129
|||||
RESULT 5
KAC_HUMAN STANDARD: PRT; 106 AA.
AC P01834;
DT 21-JUL-1986 (rel. 01, Created)
DT 21-JUL-1986 (rel. 01, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE 19 kappa chain C region.
GN IGKC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=71064023; PubMed=5489770;
RX Gotlib P.D., Cunningham B.A., Rutishauser U., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. VI. Amino
RT acid sequence of the light chain.";
RL Biochemistry 9:3155-3161(1970).
RN [2]
RP DISULFIDE BONDS.
RX MEDLINE=71064027; PubMed=4923144;
RA Gall W.E., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. X.
RT Intrachain disulfide bonds.";
RL Biochemistry 9:3188-3196(1970).
RN [3]
RP MEDLINE=72188439; PubMed=5027703;
RX Suter L., Barnikol H.U., Watanabe S., Hilschmann N.;
RT "Rule of antibody structure. The primary structure of a monoclonal
RT immunoglobulin L-chain of kappa-type, subgroup 3 (Bence-Jones protein
RT 1). IV. The complete amino acid sequence and its significance for
RT the mechanism of antibody production.";
RL Hoppe-Seyler's Z. Physiol. Chem. 353:189-208(1972).
RN [4]
RP MEDLINE=81042304; PubMed=6775818;
RX Hieber P.A., Max E.E., Seidman J.G., Maizel J.V. Jr., Leder P.;
RT "Cloned human and mouse immunoglobulin constant and J region
RT genes conserve homology in functional segments.";
RL Cell 22:197-207(1980).
RN [5]
RP MEDLINE=72188439; PubMed=5027703;
RX Suter L., Barnikol H.U., Watanabe S., Hilschmann N.;
RT "Rule of antibody structure. The primary structure of a monoclonal
RT immunoglobulin L-chain of kappa-type, subgroup 3 (Bence-Jones protein
RT 1). IV. The complete amino acid sequence and its significance for
RT the mechanism of antibody production.";
RL Hoppe-Seyler's Z. Physiol. Chem. 353:189-208(1972).
RN [6]
RP MEDLINE=68242259; PubMed=5586923;
RX Hilschmann N.;
RT "The complete amino acid sequence of Bence Jones protein Cum (kappa-
RT type)."
RL Hoppe-Seyler's Z. Physiol. Chem. 348:1718-1722(1967).
RN [7]
RP MEDLINE=69234734; PubMed=4893682;
RX Titani K., Shinoda T., Putnam F.W.;
RT "The amino acid sequence of a kappa type Bence-Jones protein. 3. The
RT complete sequence and the location of the disulfide bridges.";
RL J. Biol. Chem. 244:3550-3560(1969).
RN [8]
RP MEDLINE=70201507; PubMed=5447531;
RX Kohler H., Shlimizu A., Paul C., Putnam F.W.;
RT "Macroglobulin structure: variable sequence of light and heavy
RT chains.";

```

```

RL Science 169:56-59(1970).
CC -1- MISCELLANEOUS: THE EU SEQUENCE HAS THE INV (3) ALLOTYPIC MARKER,
CC 45-ALA & 83-VAL. THE ROY SEQUENCE HAS THE INV (1,2) ALLOTYPIC
CC MARKER, 45-ALA AND 83-LEU.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J00241; AAA58989.1;
CC EMBL: V00557; CAN23823.1;
CC PIR: A02116; K3H0.
CC HSSP: P01842; 7FAB.
CC Genew: HGNC:5716; IGKC.
CC MIM: 147200;
CC InterPro: IPR003006; Ig_MHC.
CC InterPro: IPR003597; Ig_c1.
CC SMART: SM00407; IGH1.1.
CC PROSITE: PS00290; IG_MHC; 1.
CC Immunoglobulin domain; Immunoglobulin C region.
CC FT NON TER 1 1
CC FT DISULFID 26 86 INTERCHAIN (WITH A HEAVY CHAIN).
CC FT DISULFID 106 106 V -> L (IN INV(1,2) MARKER).
CC FT VARIANT 83 83 /FTID=VAR_003897.
CC FT CONFLICT 14 14 D -> N (IN REF. 7 AND 8).
CC FT CONFLICT 57 57 E -> Q (IN REF. 5 AND 6).
CC SQ SEQUENCE 106 AA; 11609 MW; 51984D1FDD372CE8 CRC64;

Alignment Scores:
Pred. No.: 7 3e-33 Length: 106
Score: 548.00 Matches: 106
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 31.42% Indels: 0
DB: 1 Gaps: 0

US-09-806-276a-3 (1-962) x KAC_HUMAN (1-106)
QY 438 ACTGTGGCTGCACCATCTGTTTCATCTTCCGCCCATCTGATGACACTGAATTCGGA 497
|||||
DB 1 ThrValAlaAlaProSerValPheIlePheProProSerAspGluGlnLeuIysSerI 20
QY 498 ACTGCGCTGTGTGGTGGCGGCTGATTAACCTCTATCCAGAGAGGCCAAAGTACAGTGG 557
|||||
DB 21 ThrAlaSerValValCysLeuLeuAsnAsnPhenylProArgGluAlaIysValGlnTrp 40
QY 558 AAGGTGATTAAGCGCCCTCCATCGGTAACTCCAGAGAGTGTACAGACAGACAGAC 617
|||||
DB 41 LysValAlaPaspAsnAlaLeuGlnSerGlyAsnSerGlnGlnSerValThrGlnAspSer 60
QY 618 AAGACAGACAGCTTAAGCCCTCAGACAGACCCCTGAGCGCTGAGCAAGACAGACTACGAGAA 677
|||||
DB 61 LysAspSerThrThrSerLeuSerSerThrLeuThrLeuSerIysAlaSpIyrGluIys 80
QY 678 CACAAGCTACAGGCTCGAAGTCACCAACCATCAGGCGCTGAGCTGCGCTCAAAAGAGC 737
|||||
DB 81 HisIysValIyrrAlaIyGluValThrHisGlnGlyLeuSerSerProValThrIysSer 100
QY 738 TTCACACAGGAGAGAGTGT 755
|||||
DB 101 PheAsnArgGlyGluIys 106

RESULT 6
KAC_HUMAN STANDARD: PRT; 115 AA.
AC P04433;
DT 13-AUG-1987 (rel. 05, Created)

```

SEQUENCE 115 AA; 125/5 MW; 2DE47CDA3A17D555 CRC64;

```

QY      354 GAAGATGGGACCTTATATACGTACAGCAATTTTACTACTCCG 398
Db      101 GUAASPHEALValIrrytrCysGInGInArgSerAsntrpPro 115
                                     |||
RESULT 7
KV3J_HUMAN
ID KV3J_HUMAN STANDARD: PRT: 116 AA.
AC P04434;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG kappa chain V-II region VH precursor (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RX MEDLINE=85087932; PubMed=6440122;
RA Pech M., Zachau H.G.;
RT "Immunoglobulin genes of different subgroups are interdigitated
RT within the VK locus.";
RL Nucleic Acids Res. 12:9229-9236(1984).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaborati
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on t
CC use by non-profit institutions as long as its content is in no w
CC modified and this statement is not removed. Usage by and for commerci
CC entities requires a license agreement (See http://www.isb-sib.ch/announc
or send an email to license@isb-sib.ch).
-----
DR EMBL: X02725; ; NOT_ANNOTATED_CDS.
DR PIR: A01901; K3HUVH.
DR HSP: P80362; IWTI.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; Ig; 1.
DR SMART; SM00406; Ig; 1.
KW Immunoglobulin V region; Signal.
FT SIGNAL 1 20
FT CHAIN 21 >116 IG KAPPA CHAIN V-II REGION VH.
FT DOMAIN 21 43 FRAMEWORK-1.
FT DOMAIN 44 55 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 56 70 FRAMEWORK-2.
FT DOMAIN 71 77 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 78 109 FRAMEWORK-3.
FT DOMAIN 110 116 COMPLEMENTARITY-DETERMINING-3.
FT DISULFID 43 109 BY SIMILARITY.
FT NON_TER 116 116
SQ SEQUENCE 116 AA; 12757 MW; 51CD55BA5B21929 CRC64;
-----
Alignment Scores:
Pred. No.: 1.21e-28 Length: 116
Percent Similarity: 490.50 Matches: 98
Percent Similarity: 87.07% Conservative: 3
Best Local Similarity: 84.48% Mismatches: 14
Query Match: 28.12% Indels: 1
DB: 1 Gaps: 1
US-09-806-276A-3 (1-962) x KV3J_HUMAN (1-116)
QY 54 ANGAGAGCCCAAGCAGCTTCCTTCGCGCTACTGTGGTCCAGATACCAAGC 113
Db 1 MetGALAPROalagInleuNeuHeuLeuNtrPleuPROaSPThrThArg 20
114 GAATGTGTGACACAGCTTCACCAACCCCTGTCTTTGTGTCCAGGGGAAAGCACC 173
21 GtulleValMetThrGlnSerProProrHrHeuSerLeuSerProGlyGluArgValThr 40
174 CTTCTCTGAGGAGGCCAGTCAGAGCTT---AGCAGCTACTTACCTGTGACCAAGCAA 230

```

Dd	81	ProaspaphealavalTYRYCYSGInGlnTfYgIySerSerProGlnThrPhesly	100
OY	411	CAGGGACCGCTGCAGATCAACGA	437
Dd	101	GInglyserlysValGlnleltySarg	109
<hr/>			
RESULT 9	KV3D_HUMAN	STANDARD:	PRT; 109 AA.
AC	P01622;		
DT	21-JUL-1986 (rel. 01, Created)		
DT	21-JUL-1986 (rel. 01, Last sequence update)		
DT	15-JUL-1999 (rel. 38, Last annotation update)		
DE	Ig kappa chain V-IIJ region TI.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE:		
RF	MEDLINE=72188439; PubMed=5027703;		
RA	Suter L., Barnikol H.U., Wetanabe S., Hilschmann N.;		
RT	"Rule of antibody structure". The primary structure of a monoclonal immunoglobulin L-chain of kappa-type, subgroup 3 (Bence-Jones protein T1). IV. The complete amino acid sequence and its significance for the mechanism of antibody production.";		
RL	Hoppe-Seyler's Z. Physiol. Chem. 353:189-208(1972).		
CC	-1- MISCELLANEOUS: THE C REGION OF THIS CHAIN HAS THE INV (3) MARKER.		
DR	PIR: A01895; K3HUTL.		
DR	HSSP; P80362; 1MTL.		
DR	InterPro: IPR003006; Ig_MHC.		
DR	InterPro: IPR003596; I9_V.		
DR	fAm; PF00047; Ig_1.		
DR	SMART; SM00406; Igv_1.		
FT	Immunoglobulin V region; Bence-Jones protein.		
FT	DISUFID 23 89 BY SIMILARITY.		
FT	NON_TER 109 109		
SQ	SEQUENCE 109 AA; 11788 MW; 8C35058CD07749BC CRC64;		
<hr/>			
Alignment Scores:			
Pred. No.:		1.08e-27	Length: 109
Score:		477.50	Matches: 94
Percent Similarity:		93.58%	Conservative: 8
Best Local Similarity:		86.24%	Mismatches: 6
Query Match:		27.38%	Indels: 1
DB:		1	Gaps: 1
<hr/>			
US-09-806-276A-3 (1-962) x KV3D_HUMAN (1-109)			
OY	114	GAATTGTGTTGACACAGTCCTCCAGCACCCCTGTCTTGCTCCAGGAGAAAGGCCACC	173
Dd	1	GlnleValleuthrGlnserProGIYThrLeusertLeusertProGIYGLuArgAlatThr	20
OY	174	CTCTCTGCAGAGGGCCACTCAGA GTTTAGC---AGCTACTTAGCCTGGTACCAAGAAA	230
Dd	21	LeusercysArgalaSerGlnserValSeranserPheleuAlatPYrgInGlnlys	
OY	231	CCTGGCCAGGCTCCACGCTCTCATCTATGATGCATCCAACAGGAGCCATCGCATCCA	290
Dd	41	ProGIYglnAlaProAlaGleulelleTyValAlaSerSerArgAlaThnGyllePro	60
OY	291	CCCAGGTTCATGTGCGACTGGGCTCTGGGAGAGACTTCACCTCACCATCAGCAGACTGGAG	350
Dd	61	AsparghneserGlyserGlyserGlyThrAspPheThrLeuThrIleSerarGleuglu	80
OY	351	CCCGAAGATGTGCACCTTATTACTGTCAGCAATATTATTACTACACCGGTACA CTTTGGC	410
Dd	81	ProGlnAspPheAlaValTYRYCYSGInGlnTfYgIySerSerProSerThrPhesly	100
OY	411	CAGGGACCGCTGCAGATCAACGA	437
.....			

OY		114	GAAATTTGTTGGACACAGTCTCCAGGCACCCTGTCTTGCTTCACAGGGGAAGAAGCACACC	17
Db		1	GlutValLeuThrIngrInsrProGIlythrIleSerLeuSerProGIyGIAlarAlaThr	20
OY		174	CTCCTCCGACAGGGCCCACTCAGAGTGTTAACAGC---TACTTAGCCTGGTACCAACAGAA	230
Db		21	IleuSerIysArAlaSerInservAlaSerSergIyTyrlEuGLyTPryPcInGlnLys	40
OY		231	CCTGGCCAGGCTCCCGAGGGCTCCCATCTMTGATGTGATCCACAGGCGCCATGSCATCCCA	290
Db		41	ProGlyImlalrrokrArgLeuIleuIleTyrgIylAlaSerSerArGAlaThrGIylePro	60
OY		291	CCAGGTCAGTGGCAGTGGGTCTGTGGAGACAGACTCACCTCCACATCAGCAGAGCTTGAG	350
Db		61	AsparIpheserSergIySergIySergIyThrAspMetThrIleuThrIleSerArGLeuGlu	80
OY		351	CCCGAAGATGTGGCACTATTATTACGTACGACAATATTTTACTACTCCGTACACTTTGGC	410
Db		81	ProGlnAspHeAlaValIlyTrtyrCysGlnGlnIntyrgIyserLeuGIyArGThrPhgGly	100
OY		411	CAGGGACACAGCGTGGAGATCAAACGA	437
Db		101	GlnGIyThrlYsValGlnIleLysArg	109
RESULT 11				
KV3A_HUMAN				
ID KV3A_HUMAN	STANDARD:	PRT:	108 AA.	
NC P01619;				

QY	174	CTTCCCTGCGAGGGGCACATCAGAGTGTTTACC---	AGCTACTTAAGCCTGTGTACCAACAGAAA	230
Dd	21	LeuSerfyrAlgaIaaSerInSerLeuSerGlyaaNryrLeuAlaIatPrTyGIingInLys	40	
OY	231	CCTGGCCAGAGCTCCAGAGCTCTCATCTTCATWGAATGCATCCAACAGAGGCCACATGGCATGCCA	290	
Dd	41	ProGlyIgmIaIproArIgreuleuemetYrGLyValIseSerAlgaIathngILylePro	60	
OY	291	CCCAGGTTTCAGTGGCGAGTGGGTCTGGGACAGACTTCACCTTCACCAATCAGACAGACTGGAG	350	
Dd	61	AspaTrgheserIylserIylserGIalaaSphetrhrLeuThrIleSerArgIeue***	80	
OY	351	CCCCAGATGTGGCTACTTTATTACTGTGCAGCAATTTTTACTACTCCGTCACACTTTTGGC	410	
Dd	81	ProGlnasphenelaaIalyrrTyrgsingIntnryGIylserSerProphetrhrPhagily	100	
OY	411	CAGGGGACACAGGCGTCGAGATCAA	434	
Dd	101	GInglyserIylserIshleuGluIleLlys	108	
<hr/>				
RESULT 12				
ID	KV3G_HUMAN			
AC	P04206;	STANDARD;	PRT;	109 AA.
DT	20-MAR-1987 (Rel. 04, Created)			
DT	20-MAR-1987 (Rel. 04, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, last annotation update)			
DE	Ig kappa chain V-IIJ region CDR1 (Rheumatoid factor).			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
IN	[1]			

RP SEQUENCE.  
 RA MEDLINE=86230578; PubMed=3086710;  
 RX Newkirk M., Chen P.P., Carson D.A., Posnett D., Capra J.D.;  
 RT "Amino acid sequence of a light chain variable region of a human  
 RT rheumatoid factor of the Wa idiotype group, in part predicted by its  
 RT reactivity with anti-peptide antibodies.";  
 RL Mol. Immunol. 23:239-244(1986).  
 DR PIR: A01893; KAHUGO.  
 DR HSSP: P80362; 1MTL.  
 DR InterPro: IPR003006; 1g\_MHC.  
 DR InterPro: IPR003596; 1g\_V.  
 DR Pfam: PF00047; 1g; 1.  
 DR SMART: SM00406; IGV: 1.  
 DR Immunoglobulin V region.  
 KW DISULFID 23 89  
 FT NON\_TER 109 109  
 SQ SEQUENCE 109 AA; 11850 MW; 9349A5BD93588B6 CRC64;  
 BY SIMILARITY.

Alignment Scores:  
 Pred. No.: 6,17e-26 Length: 109  
 Score: 453.50 Matches: 89  
 Percent Similarity: 89.91% Conservative: 9  
 Best Local Similarity: 81.65% Mismatches: 10  
 Query Match: 26.00% Indels: 1  
 Gaps: 1

US-09-806-276a-3 (1-962) x KV3G\_HUMAN (1-109)

OY 114 GAATGTGTGTACACAGCTCCAGCCAGCCGCTGTCTGTCTCCAGGGAAGCCACC 173  
 |||||  
 Db 1 GlllvalleulheulheulserprogllythrleulserprogllylalaThr 20  
 OY 174 CTCTCCTGCAGGCGCAGT---CAGAGTGTAGACACTTACCTGTGTACACAGCAAA 230  
 |||||  
 Db 21 LeulserCysargAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 40  
 OY 231 CCGGCGCAGGCTCCGCGCTCCTCATGTATGATGCATCCAGGCGCAGCTGCATCCCA 290  
 |||||  
 Db 41 ProgllylAlaProAlaProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 60  
 OY 291 CCGGAGTTCAGTGGCAGTGGGCTGTGGAGACAGCTTCACTTCACATCAGACAGCTGAG 350  
 |||||  
 Db 61 AspArgPheSerIysSerIysSerIysSerIysSerIysSerIysSerIysSerIys 80  
 OY 351 CCGGAGATGTGCCTATTATTCTCTGTCAGCAATATTACTACTCCGACACTTTGGC 410  
 |||||  
 Db 81 ProGlnAspPheAlaValIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 100  
 OY 411 CAGGCGACAGGCTGCAGATCAACGA 437  
 |||||  
 Db 101 GlnGlyThrIy 109

RESULT 13  
 KY1W\_HUMAN STANDARD; PRT; 129 AA.  
 ID P04431.  
 AC 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 13-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig kappa chain V-I region Walker precursor.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=85014148; PubMed=6091049;  
 RX Klobeck H.G., Combrato G., Zachau H.G.;  
 RT "Immunoglobulin genes of the kappa light chain type from two human  
 RT lymphoid cell lines are closely related.";  
 RL Nucleic Acids Res. 12:6995-7006(1984).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
 CC -----  
 DR EMBL: X00965; CAA25477.1; ALU\_TERM.  
 DR PIR: A01893; KIHOMK.  
 DR HSSP: P01607; 1REI.  
 DR InterPro: IPR003006; 1g\_MHC.  
 DR InterPro: IPR003596; 1g\_V.  
 DR Pfam: PF00047; 1g; 1.  
 DR SMART: SM00406; IGV: 1.  
 DR Immunoglobulin V region: Signal.  
 KW SIGNAL 1 22  
 FT CHAIN 23 129  
 FT DOMAIN 23 45  
 FT DOMAIN 46 56  
 FT DOMAIN 57 71  
 FT DOMAIN 72 78  
 FT DOMAIN 79 110  
 FT DOMAIN 111 119  
 FT DOMAIN 120 129  
 FT DISULFID 45 110  
 FT NON\_TER 129 129  
 SQ SEQUENCE 129 AA; 14069 MW; F941FA07DA4FC2P9 CRC64;  
 BY SIMILARITY.

Alignment Scores:  
 Pred. No.: 3,13e-25 Length: 129  
 Score: 844.00 Matches: 86  
 Percent Similarity: 81.10% Conservative: 17  
 Best Local Similarity: 67.72% Mismatches: 24  
 Query Match: 25.46% Indels: 0  
 Gaps: 0

US-09-806-276a-3 (1-962) x KY1W\_HUMAN (1-129)

OY 54 ATGGAGCCCGCAGCTCAGCT 113  
 |||||  
 Db 3 MetArgValProAlaGlnLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 22  
 OY 114 GAATGTGTGTACACAGCTCCAGCCAGCCGCTGTCTGTCTCCAGGGAAGCCACC 173  
 |||||  
 Db 23 AspIleGlnMetThrGlnSerProSerSerSerSerSerSerSerSerSerSerSerSer 42  
 OY 174 CTCTCCTGCAGGCGCAGTGCAGAGTGTAGACACTTACCTGTGTACACAGCAAACT 233  
 |||||  
 Db 43 IleThrCysArgAlaSerGlnSerIleSerIleSerIleSerIleSerIleSerIleSer 62  
 OY 234 GCGCAGGCTCCCGAGCTCCTCATGTATGATGCATCAACAGGCGCAGCTGCATCCCA 293  
 |||||  
 Db 63 GlyIyAlaProIyIyLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 82  
 OY 294 AGGTTCAGTGGCAGTGGGCTGTGGAGACAGCTTCACTTCACATCAGACAGCTGAGCC 353  
 |||||  
 Db 83 ArgPheSerIySerIySerIySerIyThrAspPheThrIleSerSerLeuGlnPro 102  
 OY 354 GAGATGTGCCTATTATTCTCTGTCAGCAATATTACTACTCCGACACTTTGGCAG 413  
 |||||  
 Db 103 GlnAspSerAlaThrIy 122  
 OY 414 GGGACGAGGCTGCAGATCAAA 434  
 |||||  
 Db 123 GlyThrArgLeuGlnIleIyIy 129

RESULT 14  
 KY4C\_HUMAN STANDARD; PRT; 134 AA.  
 ID P06314.  
 AC 01-JAN-1988 (Rel. 06, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Ig kappa chain V-IV region B17 precursor.  
OS Homo sapiens (Human).  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86041854; PubMed=2997713;  
RA Marsh P., Mills F., Gould H.;  
RT "Detection of a unique human V kappa IV germline gene by a cloned  
RT cDNA probe.";  
RL Nucleic Acids Res. 13:6531-6544 (1985).  
RN [2]  
RP REVISION TO 76.  
RA Marsh P.;  
RL Submitted (OCT-1986) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration -  
CC the European Bioinformatics Institute. There are no restrictions on  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR	EMBL: 002990; CAA26733.1; -	
DR	PIR: A01905; K4HI07.	
DR	HSSP: P80362; 1WTL.	
DR	InterPro: IPR003006; IG_MHC.	
DR	InterPro: IPR003596; IG-V.	
DR	Pfam: PF00047; 1g; 1.	
DR	SMART: SM00406; IGV; 1.	
KW	Immunoglobulin V region; signal.	
FT	SIGNAL	1
FT	CHAIN	21
FT	DOMAIN	21
FT	DOMAIN	44
FT	DOMAIN	61
FT	DOMAIN	76
FT	DOMAIN	83
FT	DOMAIN	115
FT	DOMAIN	122
FT	DISULFID	133
FT	NON_TER	134
SQ	SEQUENCE	134 AA; 14966 MW; 6413A22EFD0738832 CRC64;

Alignment Scores:	
Pred. No.:	4.4e-25
Score:	442.00
Percent Similarity:	78.36%
Best Local Similarity:	61.54%
Query Match:	25.34%
Length:	134
Matches:	83
Conservative:	22
Mismatches:	23
Indels:	6

US-09-806-276A-3 (1-962) x KV4C\_HUMAN (1-134)

QY	54	ATGGAAACCCACACCTAGCTTCCTTCCTCCGTCAGTCCTGGCTCCAGATACCAACCCGGA	113
Db	1	MetValLeuLeuInrInrGlnValPheIleSerLeuLeuTrpPileSerGlyAlaTygIle	20
QY	114	GAATTTGTGTGCACACATGCTCCACCAACCTGTCTTGTCTCCAGGGGAAGACCAAC	173
Db	21	AspIleValMetThrGlnSerProAspSerLeuAlaValSerLeuGlyGluAlaGlyAlaThr	40
QY	174	CTCTCCCGCAGGGCCGATGACGTGT-----ACAGCTACTTACGCC	215
Db	41	IleAsnGlySerSerSerIleLeuTyrSerSerAspAsnLysAsnTyrLeuAla	60
QY	216	TGTATACCAACAGAAACTGCGAGGCTCCCGAGGCCCTCATCTATGATGCATCCAAAG	275
Db	61	TrpTyrGlnGlnLysProGlyeInrProProLysLeuLeuIleTyrTrpAlaSerThrArg	80
QY	276	GCCACTGGACATCCACCCAGGTTCACTAGTGGAGTGGGTCTGGAGACAGACTTACATTCACC	335

Accession	Protein	Length
Db	Glucerylalylprolinsparpheserelyserylserglythrraspherinleuthr	100
QY	ATCAGCAGACTGAGGCCGAGAGATGTGGCACATTTATCTGCAGCAATATTTTAACT	395
Db	lleserterleuglnalacilunspalmlavalyltryrtyrsglnglnintyrryasnleu	120
QY	CCGACACTTTTGGCCAGGGACGCGCTGGAGATCAACGA	437
Db	Protriphphenelylglnelnythrrylsvalguilelystac	134

RESULT 15				
KV3F_HUMAN				
ID	KV3F_HUMAN	STANDARD;	PRT;	109 AA.
AC	P01624;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Iq kappa chain V-II region F0M.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RN	SEQUENCE.			
RX	MEDLINE=76276460; PubMed=60899;			
RA	Klapper D.G., Capra J.D.;			
RT	"The amino acid sequence of the variable regions of the light chains			
RT	from two idiotypically cross reactive Igm anti-gamma globulins.";			
RL	Ann. Immunol. (Paris) 127C:261-271(1976).			
CC	-I- GLOBULIN. THIS CHAIN WAS ISOLATED FROM AN IGM WITH ANTI-GAMMA			
	GLOBULIN ACTIVITY.			

SEQ	SEQUENCE	109 AA:	11922 MW:	62821DDCG6A8A86 CRC64:
DR	PIR; A01897; K3HDM.			
DR	HSSP; P80362; 1WTL.			
DR	InterPro; IPR003206; Ig_MHC.			
DR	InterPro; IPR003596; Ig_V.			
DR	Pfam; PF00047; Ig; 1.			
DR	SMART; SM00406; IGV; 1.			
KW	Immunoglobulin V region.			
ET	DISULFID	23	89	
ET	NON_TER	109	109	
SO	SEQUENCE	109 AA:	11922 MW:	62821DDCG6A8A86 CRC64:

BY SIMILARITY.

Alignment Scores:	
Pred. No.:	6, 55e-25
Score:	439, .50
Percent Similarity:	87.16%
Best Local Similarity:	79.82%
Query Match:	25, 20%
GB:	1
Length:	109
Matches:	89
Conservative:	8
Mismatches:	13
Indels:	1
Gaps:	1

US-09-806-276A-3 (1-962) x KV3F\_HUMAN (1-109)

[illegible]

Fri Jul 25 09:59:16 2003

us-09-806-276a-3.rsp

Page 10

QY 411 CAGGGACGAGCTGAGATCAACGA 437  
|||||  
Db 101 GInGIYtMrArGvAlGulIeIySArG 109

Search completed: July 18, 2003, 17:26:19  
Job time : 16.0259 secs



Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	1352.5	77.6	414	12	AAR13111	1B1 Igg aberrant
2	1352.5	77.6	414	12	AAR13018	1B1 Igg aberrant
3	1212	69.5	234	21	AAV92239	Human bone marrow
4	1133.5	65.0	235	21	AAV93702	The kappa chain o
5	1133.5	65.0	235	21	AAV93729	The kappa chain o
6	1131	64.9	234	21	AAO14066	Light chain prote
7	1127	64.6	234	21	AAV93708	The kappa chain o
8	1127	64.6	234	21	AAV93708	The kappa chain o
9	1116	64.0	236	22	AAV93733	Human gene 2-enco
10	1116	64.0	236	22	AAV93733	Human albumin fus
11	1114.5	63.9	233	21	AAV93704	The kappa chain o
12	1114.5	63.9	233	21	AAV93704	The kappa chain o
13	1108	63.5	236	23	AAV93721	Anti-human ALIM
14	1098	63.0	236	23	AAV93729	Anti-human ALIM
15	1084	62.2	234	15	AAV92951	Human anti-IgE MA
16	1073	61.5	234	22	AAV92951	Human immune syst
17	1055.5	60.4	238	19	AAV92951	1B1 Igg aberrant
18	1053	60.4	238	19	AAV92951	Anti-Fas humanise
19	1053	60.4	238	21	AAV92951	Humanised anti-Fa
20	1053	60.4	238	21	AAV92951	Humanised anti-Fa
21	1053	60.4	238	23	AAV92951	Humanised anti-Fa
22	1053	60.4	238	23	AAV92951	Humanised anti-Fa
23	1047	60.0	238	19	AAV92951	Humanised anti-Fa
24	1047	60.0	238	21	AAV92951	Humanised anti-Fa
25	1047	60.0	238	21	AAV92951	Humanised anti-Fa
26	1047	60.0	238	23	AAV92951	Humanised anti-Fa
27	1047	60.0	238	23	AAV92951	Humanised anti-Fa
28	1042.5	59.8	384	22	AAV92951	Human novel prote
29	1042.5	59.8	384	22	AAV92951	Human novel prote
30	1042.5	59.8	384	22	AAV92951	Human novel prote
31	1042.5	59.8	384	22	AAV92951	Human novel prote
32	1035	59.3	214	20	AAV92951	Human novel prote
33	1033	59.2	224	22	AAV92951	Human novel prote
34	1033	59.2	224	22	AAV92951	Human novel prote
35	1033	59.2	224	22	AAV92951	Human novel prote
36	1033	59.2	224	22	AAV92951	Human novel prote
37	1033	59.2	224	22	AAV92951	Human novel prote
38	1033	59.2	224	22	AAV92951	Human novel prote
39	1033	59.2	224	22	AAV92951	Human novel prote
40	1029	59.0	224	22	AAV92951	Human novel prote
41	1029	59.0	224	22	AAV92951	Human novel prote
42	1029	59.0	224	22	AAV92951	Human novel prote
43	1029	59.0	224	22	AAV92951	Human novel prote
44	1029	59.0	224	22	AAV92951	Human novel prote
45	1029	59.0	224	22	AAV92951	Human novel prote

FT		/label= variable region
FT		/note= "L/V 1"
FT	Region	131..243
FT		/label= variable region
FT		/note= "L/V 2"
FT	Region	244...345
FT		/label= constant region
PX		
PN		
XX	MO9106305-A.	
PD	16-MAY-1991.	
XX		
PF	06-NOV-1990;	90MO-US06426.
XX		
PR	07-NOV-1989;	89US-0432700.
XX		
PA	(BRIM ) BRISTOL-MYERS SQUIB.	
XX		
PI	Shuford WM, Harris LJ, Ralf HV;	
XX		
DR	WPI; 1991-163947/22.	
N-PSDB:	AAQJ1878.	
PT	Oligomeric immunoglobulin(s) with high avidity for antigen(s)	-
PT	formed by duplicating esp. variable region of light chain of IgG	
PT	class	
XX		
PS	Example 5; Fig 16; 104pp; English.	
XX		
CC	This sequence is deduced from the cDNA clone 4B9-VK15 and includes	
CC	the amino acid sequence beyond the first stop codon ("x" in the	
CC	sequence represents a nonsense codon). The clone is incomplete,	
CC	starting from the G of the ATG initiator codon, but the initial Met	
CC	is given. Antibody molecules of the invention can include one or two	
CC	aberrant light chains containing a duplicated variable region, to	
CC	produce heavier antibodies. These heavier antibodies were found to	
CC	have higher avidity than antibodies with just a single copy of the	
CC	L.V region. The antibodies can be used to treat disease, e.g.	
CC	infection by streptococcus agalactiae. They are able to pass across	
CC	the placenta	
CC	See also AAQJ1879 and AAQJ1880.	
XX		
SQ	Sequence 414 AA;	
	Alignment Scores:	
	Pred. No.: 2,54e-109	Length: 414
	Score: 1352.50	Matches: 282
	Percent Similarity: 69.08%	Conservative: 4
	Best Local Similarity: 68.12%	Mismatches: 13
	Query Match: 77.55%	Indels: 115
	DB: 12	Gaps: 2
US-09-806-276A-3 (1-962) x AAR13111 (1-414)		
QY	54 ATGGAGCCCCCAGCTCAGTTCCTCTCTCTCTGCTACTCTGTGGCTCCCGATACGCCGGA	113
DB	1 MetGlnAlaProAlaGlnLeuLeuPheLeuLeuLeuTrpLeuProAspThrThnGly	20
QY	114 GAATTGCTGTGACACAGCTCCAGGCACCCTGCTTGTCCTCCAGGGAGAAGCCACC	173
DB	21 GtllleValleutrnGrInserProAlaThrLeuSerLeuSerProGlgLuArgIaThr	40
QY	174 CTCCTCATGAGGCACATCAGATGTTAGTAGAGCTACTTAAGCTGTGTAACAACAACCT	233
DB	41 LeusercysArghAlaserserInserValGlysetrTyrlenuAlaTrpTyrGlnglndysPro	60
QY	234 GGCCAGGCTCCAGGCTCTCTCATCTATGATGATCAATCCAAGGCGCACrGGATCCACCC	293
DB	61 GlycinalaProAlaProLeuLeuIetyrAspAlasernArghIaThrGlyleProla	80
QY	294 AGCTCACTGGCAGTGGGCTCGGGAGCACTTCACTCACTCAACATGAGAGCCC	353
DB	81 ArgphIsersergIysergIythrAspPherThrLeuthrIlleserserLeuSlnpro	100

QY	354	GAAGATGGCGCACTTTATTAAGTCAG-----	380
Db	101	GIuaspPheAlaValTYrTYrCYSGlnHisArgspasnTrProProGlyAlaThrPhe	120
QY	380	-----	380
Db	121	GIyGIyGIYThrLYsValGIuIIeLYsHisThrPrgLIuIIeValIeuThrGlnSer	140
		-----	
QY	380	-----	380
Db	141	ProAlaThrLeuSerLeuSerProGlyGluArgAlaThrLeuSerCYsArgAlaSerGln	160
		-----	
QY	380	-----	380
Db	161	SerValGIySerTYrLeuAlaTrpTYrGlnGlnLYsProGlyGlnAlaProArgProIeu	180
		-----	
QY	380	-----	380
Db	181	IIeTYrAspAlaSerAsnArgAlaThrGlyIIeProAlaArgPheSerGlySerGlySer	200
		-----	
QY	381	-----CAATATTATTACTACTCCGAC	401
Db	201	GIYThrAspPheThrLeuThrIIeSerSerLeuGluProGluAspPheAlaValTYrTYr	220
		-----	
QY	402	-----ACTTTGGCGCAAGGCGACGAGCTGGAG	428
Db	221	CYSGLInHisArgspasnTrProProGlyAlaThrPheGlyGlyGlyThrLYsValGIu	240
		-----	
QY	429	ATCAAAACGAAGTGTGGCTGCACCATCTGTCTTCATCTTCCCGCATCTGATGACAGTTG	488
Db	241	IIeLYsArgThrValAlaIaIaIaProSerValPheIIePheProProSerAspLIuGlnIeu	260
		-----	
QY	489	AAATCTGGAACTGCCCTGTGTGTGGCCCTGTAATTACTTATCCAGAGAGGCCAAA	548
Db	261	LYSserGlyThrAlaSerValValCYsLeuIeuAsnAsnPheTYrProArgLIuAlaLYs	280
		-----	
QY	549	GTAACACTGGAGGTGATAAAGCCCTCCAAATGGGTAACTCCAGAGAGTGTACACAG	608
Db	281	ValGInTrpLYsValAspAsnAlaIeuGlnSerGlyAsnSerGlnIeuSerValThrGlu	300
		-----	
QY	609	CAGGACAGAGAGGACGACACTTAACAGCTCAGACGACACCTGACGCTGAGCAAGCAGAC	668
Db	301	GInAspSerLYsAspSerThrTYrSerLeuSerSerThrLeuThrLeuSerLYsAlaAsp	320
		-----	
QY	669	TACGAGAAACACAAAGCTTACGGCTCGAAGTCACCCATCAGGGCTGAGCTCGCCGTC	728
Db	321	TYrGInLYsHisLYsValTYrAlaCYsGIuValThrHisGlnIeuLYsSerProVal	340
		-----	
QY	729	ACAAGAAGCTTCAACAGGGAGAGTGTTAGAGGGAGAAGTGGCCCCACCTGGCTCCACAT	788
Db	341	ThrLYsSerPheAsnArgLYsIuCYs**ArgGInLYsCYsProHisLeuLeuSer	360
		-----	
QY	789	TTCAGCTGACCCCTCCATCTCTTGGACTTGAGCCCTTTCACAGGGAGCTTACCC	848
Db	361	SerSerLeuThrProSerHisProLeuAlaSerAspProPheSerThrGlyAspLeuPro	380
		-----	
QY	849	CTATTGGGGTCTCCAGACTCATCTTTCACCTGACCCCTCCCTCTTGGCTTTAATT	908
Db	381	LeuIeuArgSerSerSerSerSerPheThrSerProProSerSerLeuAlaIeuIle	400
		-----	
QY	909	ATGCTAATGTTGGAGAGGAATGAATAATAAGTGAATCTTT	950
Db	401	MetIeuMetLeuGInGInAsnGIn**IIeLYs**IIePhe	414
		-----	
RESULT 2			
AAR13018			
ID	AAR13018 standard; Protein: 414 AA.		
XX	AAR13018:		
XX	AC		
DT	01-AUG-1991 (first entry)		
XX			

[illegible]



```

Db      221 LeuSerSerProValThrLysSerPheAsnArgGlyGluCys 234
RESULT 4
ID      AAY93702
AAV93702 standard; Protein; 235 AA.
AC      AAV93702;
DE      03-OCT-2000 (first entry)
XX      The kappa chain of immunoglobulin clone 4.1.1.
XX      Cytoxic T-lymphocyte antigen-4; CTLA-4; antibody; immune system;
XX      hyperimmunity disorder; autoimmune disease; diabetes; graft rejection;
XX      proliferative disorder; cancer; immunodeficient disorder.
XX      Homo sapiens.
XX      MO200037504-A2.
XX      29-JUN-2000.
XX      23-DEC-1999; 99WO-US30895.
XX      23-DEC-1998; 98US-0113647.
XX      (PFTZ ) PFIZER INC.
XX      (ABGE-) ABGENIX INC.
XX      Hanson DC, Neveu MJ, Mueller EE, Hanke JH, Gilman SC, Davis CG;
XX      Corvalan JR;
XX      MPI: 2000-442647/38.
XX      N-PDB: AAA46865.
XX      Novel antibodies capable of binding cytotoxic T-lymphocyte antigen
XX      (CTLA)-4 containing specified heavy and light chain sequences, useful
XX      for treating, e.g. immune disorders -
XX      Claim 3; Fig 1A; 157pp; English.
XX      The present sequence represents a kappa chain of an antibody of the
XX      invention. The antibody is directed cytotoxic T-lymphocyte antigen
XX      (CTLA)-4. Antibodies of the invention are composed of a heavy chain
XX      variable region, comprising a modified contiguous sequence from a
XX      FRI-FR3 sequence encoded by a human VH3-33 family gene. The
XX      modifications are contained in CDR1, CDR2 and/or framework regions.
XX      The antibodies may be used to inhibit CTLA-4 and/or framework regions.
XX      Immune system to treat hyperimmunity disorders (e.g. autoimmune
XX      disease, diabetes and graft rejection) and proliferative disorders
XX      (e.g. cancer). CTLA-4 stimulatory agents may be used to up-regulate
XX      immune system to up-regulate immunodeficient disorders.
XX      Sequence 235 AA;
Alignment Scores:
Pred. No.: 2,61e-90 Length: 235
Score: 1133.50 Matches: 219
Percent Similarity: 97.02% Conservative: 9
Best Local Similarity: 93.19% Mismatches: 6
Query Match: 64.99% Indels: 1
DB: 21 Gaps: 1
US-09-806-276a-3 (1-962) x AAV93702 (1-235)
QY      54 ATGAGAGCCCGCAGCTTCTCTTCCCTGCTACTGCTGCTCCAGATACCGACGGA 113
DB      1 MetGluThrProAlaSerGlnLeuLeuPheLeuLeuLeuTrpLeuProAspThrThrgly 20
QY      114 GAAATTTGTTGACAGAGTCTCCAGCCCGTCTTGTCTCCAGGGGAAAGAGCCACC 173
DB      21 GlnIleValLeuThrGlnSerProGlyThrLeuSerLeuSerProGlyGluArgAlaThr 40

```

```

QY      174 CTCTCCAGAGGCCAGTACAGAGTGT---AGCAGCTACTAGCTGTACCAACAGAAA 230
DB      41 LeuSerCysArgAlaSerGlnSerIleSerSerSerPheLeuAlaTrpTyrGlnGlnArg 60
QY      231 CTGGCCAGGCTCCAGGCTCCTCATCTATGATGCATCCACAGAGGACACGAGCATCCCA 290
DB      61 ProGlyGlnAlaProArgLeuLeuIleTyrGlyAlaSerSerArgAlaThrGlyIlePro 80
QY      291 CCCAGTTTCAGTCGAGCGGCTGGTGGGACACACTTCTACTCTACCATCAGCAGACTGGAG 350
DB      81 AspArgPheSerGlySerGlySerGlyThrAspPheThrLeuThrIleSerArgLeuGlu 100
QY      351 CCCGAAAGATGTGGCACTTATTACTGTGCAGCAATATTTTACTACTCCGTACCTTTGGC 410
DB      101 ProGluAspPheAlaValTyrTyrCysGlnGlnIleTyrGlyThrSerProTrpPheGly 120
QY      411 CAGGGAGCCAGCGCTTGGAATCAAAACGACTGTGGCTGCACCATCTGCTCATCTTCCCG 470
DB      121 GlnGlyThrLysValGlnIleLysArgThrValAlaAlaProSerValPheIlePhePro 140
QY      471 CCATCTGATGAGCAGTTGAAATCTGGAACCTGCTGTGTGTGGCTCGATTAACCTTC 530
DB      141 ProSerAspGlnGlnLeuLysSerGlyThrAlaSerValValCysLeuLeuAsnPro 160
QY      531 TATCCAGAGAGGCCAAAGTACAGTGAAGGTGGATACGCCCTCCATCGGTAACCTCC 590
DB      161 TyrProArgGlnAlaLysValGlnTrpLysValAlaPheAsnAlaLeuGlnSerGlyAsnSer 180
QY      591 CAGGAGAGTGTGCAGAGCGAGGACAGCAAGACAGACACCTACAGCTCAGACACCTCG 650
DB      181 GlnGlnSerValThrGlnGlnAspSerLysAspSerThrTyrSerLeuSerSerThrLeu 200
QY      651 AGCTGAGCAAGAGCAGACTACGAGAAACAAAGTCTACGCCGCGAGATCCATCCATCAG 710
DB      201 ThrLeuSerLysAlaAspTyrGlnLysHisLysValTyrAlaCysGlnValThrHisGln 220
QY      711 GGCTTGAGCTGCGCCCGCCACAAAGAGCTTCAACAGGGGAGAGTGT 755
DB      221 GlyLeuSerSerProValThrLysSerPheAsnArgGlyGluCys 235
RESULT 5
AAY93729
ID      AAY93729 standard; Protein; 235 AA.
XX      AAY93729;
AC      AAY93729;
DE      03-OCT-2000 (first entry)
XX      The kappa chain of immunoglobulin clone 4.1.1.
XX      Cytoxic T-lymphocyte antigen-4; CTLA-4; antibody; immune system;
XX      hyperimmunity disorder; autoimmune disease; diabetes; graft rejection;
XX      proliferative disorder; cancer; immunodeficient disorder.
XX      Homo sapiens.
XX      Key Location/Qualifiers
XX      Peptide 1..20 /note="signal peptide"
XX      MO200037504-A2.
XX      29-JUN-2000.
XX      23-DEC-1999; 99WO-US30895.
XX      23-DEC-1998; 98US-0113647.
XX      (PFTZ ) PFIZER INC.
XX      (ABGE-) ABGENIX INC.
XX      Hanson DC, Neveu MJ, Mueller EE, Hanke JH, Gilman SC, Davis CG;

```



OY	54	ATGGAAGCCCGGACAGCTTCCTGCTCTGCTACTGCTGCTCCAGATACCAGGGA	113
Db	1	MetGluAlaProAlaGlnLeuPheLeuLeuLeuLeuTrpLeuProAspTrpThrGly	20
OY	114	GAATTGTGTGACACAGTCTCCAGCACCTGTCTTGTCTCCAGGGAAAGAGCCACC	173
Db	21	GluIleValLeuThrGlnSerProAlaThrLeuSerLeuSerProGlyGluArgAlaThr	40
OY	174	CTCCGTCGACGGCCATCTCAGAGTGTAGACAGCTACTTAACTGGTACCAAGAAACCT	233
Db	41	LeuAlaCysArgAlaSerGlnThrAlaSerArgTyrLeuAlaTrpTyrGlnGlnLysPro	60
OY	234	GGCCAGGCTCCGACAGCTCTCATCTATGATGATTCACACAGGGCCACTGGCATCCACCC	293
Db	61	GlyGlnAlaProArgLeuLeuIleTyrAspTrpSerAsnArgAlaThrGlyIleProAla	80
OY	294	AGGTTCAGTGGCAGTGGGTCTGGACACAGCTTCACTCACTACACATCCAGACAGTGAAGCC	353
Db	81	ArgPheSerIleYserGlySerGlyThrAspPheThrLeuSerIleSerLeuGlnPro	100
OY	354	GAAGATGTGCACATTATTACTGTCCAGCAATATTTTACTACCTCCGACATTTGGCCAG	413
Db	101	GluAspPheAlaValIleTyrTyrGlyGlnGlnAlaArgPheAsnTrpProTrpThrPheGlyGln	120
OY	414	GGACACAGCTGGAGATCAAAACGAACTGTGGCTGCACATGTCTTCACTTCCCGGCA	473
Db	121	GlyThrLysValGlnPheLysArgThrValAlaAlaAlaProSerValPheIlePheProPro	140
OY	474	TCGTAGACAGCTTGAATCTGGAACTGGCTGTGTGTGTGTGGCTGCTGATTAATCTTAT	533
Db	141	SerAspGlnGlnLeuLysSerGlyTrpAlaSerValValCysLeuLeuAsnAsnPheTyr	160
OY	534	CCGACGAGGCGCAAACTACGTGGAAGTGGATTAACGCCCTGCATCGGGTAACTCCAG	593
Db	161	ProArgGluAlaLysValGlnTrpLysValaAspAlaLeuGlnSerIleLysAsnSerGln	180
OY	594	GAGAGTGCACAGACGAGACAGCAAGACAGCACCTACAGGCTTACGACGACCTGTAGC	653
Db	181	GluSerValThrGlnGlnAspSerLysAspSerThrTyrSerLeuSerSerThrLeuThr	200
OY	654	CTGACCAAAAGCAGCTACGAGAAACACAAAGTCTACGGCTGGAAATACCCATCAGGCG	713
Db	201	LeuSerLysAlaAspTyrGluYshHisLysValIleTyrAlaCysGluValThrHisGlnIle	220
OY	714	CTGAGCTGGCCGCTGACAAAGAGCTTCAACAGGGGAGAGTGT	755
Db	221	LeuSerSerProValThrLysSerPheAsnArgGlyGlnCys	234
RESULT 7			
ID	AAV93708	standard; Protein: 234 AA.	
XX	AAV93708;		
AC	AAV93708;		
DT	03-OCT-2000	(first entry)	
DE	The kappa chain of immunoglobulin clone 6.1.1.		
XX			
KW	Cytotoxic T-lymphocyte antigen-4; CTLA-4; antibody; immune system; hyperimmunity disorder; autoimmune disease; diabetes; graft rejection; proliferative disorder; cancer; immunodeficient disorder.		
XX			
OS	Homo sapiens.		
PN	WO200037504-A2.		
PD	29-JUN-2000.		
XX	23-DEC-1999;	99MO-US30895.	
PF	23-DEC-1998;	98US-0113647.	
XX	(PF12 ) PFIZER INC.		

PA	(ABGE-) ABGENIX INC.
XX	
PI	Hanson DC, Neveu MJ, Mueller EE, Hanke JH, Gilman SC, Davis CG;
PI	Corvalan JR;
XX	
DR	WPI: 2000-442647/38.
DR	N-PSDB; AAA46871.
XX	
PT	Novel antibodies capable of binding cytotoxic T-lymphocyte antigen
PT	(CTLA-4 containing specified heavy and light chain sequences, useful
XX	for treating, e.g. immune disorders -
PS	
XX	Claim 3; Fig 1D; 157pp; English.
CC	
CC	The present sequence represents a kappa chain of an antibody of the
CC	invention. The antibody is directed cytotoxic T-lymphocyte antigen
CC	(CTLA)-4. Antibodies of the invention are composed of a heavy chain
CC	(CTLA)-4 variable region, comprising a modified contiguous sequence from a
CC	FR1-FR3 sequence encoded by a human VHS-33 family gene. The
CC	modifications are contained in CDR1, CDR2 and/or framework regions.
CC	The antibodies may be used to inhibit CTLA-4 and down-regulate the
CC	immune system to treat hyperimmunity disorders (e.g. autoimmune
CC	disease, diabetes and graft rejection) and proliferative disorders
CC	(e.g. cancer). CTLA-4 stimulatory agents may be used to up-regulate
XX	immune system to up-regulate immunodeficient disorders.
XX	
Sequence	234 Aa;
50	

Alignment Scores:	
Pred. No.:	9..59e-90
Score:	1127.00
Percent Similarity:	95.73%
Best Local Similarity:	92.74%
Query Match:	64.62%
DB:	21
	Gaps: 0
	Mismatches: 10
	Conservative: 7
	Matches: 217
	Length: 234

US-09-806-276A-3 (1-962) x AAY93708 (1-234)

QY	54	ATGGAAAGCCCAAGCTCATCTTCCTTCCTCCGTAACCTTGGCTCCCAAGATACCAACCGGA	113
Db	1	MetglntrhrproalaglnleuileupheleuileuileutrpheleproaspthrThrGly	20
QY	114	GAATTTGTTGGACACAGTCTCCAGGCCACCTGTCTTTGTCTCCAGGGGAAGAGCCACC	173
Db	21	GlnllevalleuutrnrginsenrproglythrIleuserIeuserproglyIleargalaThr	40
QY	174	CTTCTCGAGAGGCGACATCGAGTGTATTACACTACTTAAGCTGGTACCAACAGAAACT	233
Db	41	IeuserIysmrghalaserIserIserIleuIalatrpyrGlnGlnLyspro	60
QY	234	GGCCAGGCTCCCAAGGCTCCTCATATGATATGATCCAAAGGCGACACTGGCATCCACC	293
Db	61	GlyIlelnalaproarprProleuIleuIryGlyIalaserIserIargalathrnGlyIleproasp	80
QY	294	AGGTTGATGGGAGTGGGTCTGGGACAGACTTCACTCTCACCATACAGACACTGGAGCCC	353
Db	81	ArgpheserIyseryIserIyserIylthraspethrIleutrnIleSerIargleugIupro	100
QY	354	GAAGATCGGCACATTATTAATCTGTACAGAAATATTTCATCCGCTGACATTTTGGCCAG	413
Db	101	GlnasphelaIalavalryIrcysIeInIlnryGlyIleSerProphethrIpnegIypro	120
QY	414	GGGACCAAGGTTGGAAATCAACGAACCTGGCTCCACCATCTGTTTCATCTTCCGGCA	473
Db	121	GlyThrIlyIysvalAspIleIysarIghrValAlaIalaproserIalpheIleIhepPropro	140
QY	474	TCTGATGACAGTGAATAATGTGAACCTCCCTTGTGTGGCTGGCGAAATAACTTAT	533
Db	141	SerAspIglIugIlnleuIyseryIynhrIalaserIalValIcIysIeueuIasnIpnethr	160
QY	534	CCAGAGAGGCCAAAGTACACTGGAGAGGTGATATACGCCCTCCCAATCGGATTAACCTCAG	593
Db	161	ProargIleIalIysvalIcIlntrIplysvalAspIasnIalIeueuIinseryIysnserIgin	180

Percent Similarity:	95.73%	Conservative:	7
Best Local Similarity:	92.74%	Mismatches:	10
Query Match:	64.62%	Indels:	0
DB:	21	Gaps:	0
US-09-806-276A-3 (1-962) x AAY93733 (1-234)			
QY 54	ATNGAGGCCGACGCTCACCTTCTCTCCCTGCTACTGTGGCTCCAGATACACGGGA	113	
Db 1	MeuGILthProAlaISgInLeuPheLeuLeuLeuLeuLeuPheProAspThrIle	20	
QY 114	GAATTTGTGTACACAGCTCTCCACACCCCTGCTTTGTCTCCAGGGGAAGACCGAC	173	
Db 21	GIuILleValLeuIthGInSerProDILthLeuSerLeuSerProDILyGluAlaThr	40	
QY 174	CTCTCCGACGGCCGCTGACAGTGTGTAGCGACTTACCTGGTACCAAGAAACCT	233	
Db 41	LeuSerCysArgAlaSerGInSerValSerSerGlyLeuAlaTrpGInGInIysPro	60	
QY 234	GGCCAGGCTCCAGGCTCTCTATCTATATGATCATCCAAAGGGCCAGTGGCATCCACCC	293	
Db 61	GIyGInAlaIProArProLeuIleIyGILyAlaISerSerIArgAlaThrGILyIleProAsp	80	
QY 294	AGGTTGAGTGGCGAGTGGTCTGGGACAGACTTCACTCTCAACATCAGACACTGGAGCC	353	
Db 81	ArgPheSerIlySerIySerGILyThrAspThrIleSerArgLeuGILyPro	100	
QY 354	GAAGATGTGGCACTTTATTTACTGTACAGCATATTTTACTACCTGGTACACTTTTGGCCAG	413	
Db 101	GIuAspPheAlaValIyTrpCysGInGInIyGILyIleSerProPheThrPheGILyPro	120	
QY 414	GGGACCAAGCGGCGAGATCAACCAACTGTGGCTGCACCATGTCTTCATCTTCCGCCA	473	
Db 121	GIyThrIyValAspIleIyAspThrValAlaAlaProSerValIlePheIlePheProPro	140	
QY 474	TCTGATAGACAGTTGAATCTGTGACACTGCCCTGTGTGTGCTGGCGGAAATCTGTAT	533	
Db 141	SerAspGInGInLeuIySerGILyThrAlaSerValIyCysLeuLeuAspAspPheIy	160	
QY 534	CCCAAGAGGGCCAAAGATACAGTGGAGGATGAATAGGCCCTCCAAATGGGTTACTCCAC	593	
Db 161	ProArgGILyAlaIyValIyGInTrpIyValAspAspAlaLeuGInSerGILyAsnSerGIn	180	
QY 594	GAGACTGTCAAGACGACGACAGCAACAGACAGCACTTACAGCTTCAGACAGACCTGACG	653	
Db 181	GIuSerValIthGInGInAspSerIyAspSerIthIySerLeuSerSerIthLeuThr	200	
QY 654	CTGACCAAGGAGACTCGAGAAACCAAGTGTACGCTGGCAAGTCACCCATCAGGGC	713	
Db 201	LeuSerIySAlaAspTrpGILyIyHisIyValIyAlaCysGILyValIthHisGILy	220	
QY 714	CTGAGCTCGCCGTCACAAAGCTTCAACAGGGGAGAGTGT	755	
Db 221	LeuSerSerProValIthIySerPheAsnArgIyGILyCys	234	
RESULT 9			
ID AAG71272	standard. Protein: 236 AA.		
XX AAG71272:			
XX 30-JUL-2001	(first entry)		
XX Human gene 2-encoded secreted protein HTSER67, SEQ ID NO:121.			
XX Human: secreted protein: proliferative disorder: cancer: chromosome 1;			
KW foetal abnormality; developmental abnormality; hematopoietic disorder;			
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;			
KW inflammation; allergy; neurological disorder; Alzheimer's disease;			
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;			
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;			
KW cardiovascular disorder; angiogenic disorder; kidney disorder;			
KW gastrointestinal disorder; pregnancy-related disorder; tumour;			







```
Conservative: 6
Mismatches: 1
Indels: 1
Gaps: 1
```

1

FT		/note= "signal peptide"
XX		
PN	WO200037504-A2.	
PD	29-JUN-2000.	
PF		
XX		
PR	23-DEC-1999; 99WO-US30895.	
XX		
PR	23-DEC-1998; 98US-0113647.	
XX		
PA	(PFI2 ) PFIZER INC.	
PA	(ABGE-) ABGENIX INC.	
PI	Hanson DC, Neveu MJ, Mueller EE, Hanke JH, Gilman SC, Davis CG,	
PI	Corvalan JR;	
DR	MPL; 2000-442647/38.	
DR	N-PDSB; AAA46895.	
XX		
PT	Novel antibodies capable of binding cytotoxic T-lymphocyte antigen	
PT	(CTLA)-4 containing specified heavy and light chain sequences, useful	
XX	for treating, e.g. immune disorders -	
PS	Claim 3; Fig 22k; 157pp; English.	
XX		
CC	The present sequence represents a kappa chain of an antibody of the	
CC	invention. The antibody is directed cytotoxic T-lymphocyte antigen	
CC	(CTLA)-4. Antibodies of the invention are composed of a heavy chain	
CC	variable region, comprising a modified contiguous sequence from a	
CC	FRI-FR3 sequence encoded by a human VH3-33 family gene. The	
CC	modifications are contained in CDP1, CDR2 and/or framework regions.	
CC	The antibodies may be used to inhibit CTLA-4 and down-regulate the	
CC	immune system to treat hyperimmunity disorders (e.g. autoimmune	
CC	disease, diabetes and graft rejection) and proliferative disorders	
CC	(e.g. cancer). CTLA-4 stimulatory agents may be used to up-regulate	
XX	immune system co up-regulate immunodeficient disorders.	
SO	Sequence 233 AA:	
 Alignment Scores:		
Pred. No.:	1.17e+88	
Score:	1114.50	Length: 233
Percent Similarity:	95.30%	Matches: 217
Best Local Similarity:	92.74%	Conservative: 6
Query Match:	63.90%	Mismatches: 10
DB:	21	Indels: 1
	Gaps: 1	
 US-09-806-276A-3 (1-962) x AAY93731 (1-233)		
QY	54 ATGGAAGCCCGACGGCTTCCTTCTCCTCGCTACTGTGGATCCAGATCAACCGGA	11
Dd		
	1 MetGIutmrPrvIaaGlnLeuLeuPhleLeuLeuLeuTrpLeuProAsphThrIrgy	20

FT	/note= "signal peptide"
XX	
PN	WO200037504-A2.
PD	
PB	29-JUN-2000.
PF	
PR	23-DEC-1999; 99WO-US30895.
XX	
PR	23-DEC-1998; 98US-0113647.
PA	(PFI2 ) PFIZER INC.
PA	(ABGE-) ABGENIX INC.
PI	Hanson DC, Neveu MJ, Mueller EE, Hanke JH, Gilman SC, Davis CG;
PI	Corvalan JR.
DR	NPI; 2000-442647/38.
DR	N-PDSB; AAA46895.
PT	
PT	Novel antibodies capable of binding cytotoxic T-lymphocyte antigen
PT	(CTLA)-4 containing specified heavy and light chain sequences, useful
XX	for treating, e.g., immune disorders -
XX	
PS	Claim 3; Fig 22K; 157pp; English.
CC	
CC	The present sequence represents a kappa chain of an antibody of the
CC	invention. The antibody is directed cytotoxic T-lymphocyte antigen
CC	(CTLA)-4. Antibodies of the invention are composed of a heavy chain
CC	variable region, comprising a modified contiguous sequence from a
CC	FRI-FR3 sequence encoded by a human VH3-33 family gene. The
CC	modifications are contained in CDR1, CDR2 and/or framework regions.
CC	The antibodies may be used to inhibit CTLA-4 and down-regulate the
CC	immune system to treat hyperimmunity disorders (e.g., autoimmune
CC	disease, diabetes and graft rejection) and proliferative disorders
CC	(e.g., cancer). CTLA-4 stimulatory agents may be used to up-regulate
CC	immune system to up-regulate immunodeficient disorders.
XX	
Sequence	233 AA:
Alignment Scores:	
Pred. No.:	1.17e-88 Length: 233
Score:	1114.50 Matches: 217
Percent Similarity:	95.30% Conservative: 6
Best Local Similarity:	92.74% Mismatches: 10
Query Match:	63.90% Indels: 1
DB:	21 Gaps: 1
US-09-806-2764-3 (1-962) x AAY93731 (1-233)	
OY	54 ATGGAAGCCCGCAGCTGCAGCTTCCTCCCTCCGTACTGTGGCTCCAGATACCGAGA 11
Db	
	1 MetGIuThrProAlaGlnLeuLeuPheLeuLeuLeuTrpLeuProAspThrIleGly 20
OY	114 GAATTTGTTGTGACACAGTCTCCAGCACCCCTGCTTGTGCTCCAGGGAAAGACCACC 173
Db	
	21 GluIleValLeuThrGlnSerProGlyThrLeuSerLeuSerProGlyGluArgAlaThr 40
OY	174 CTCTCTGCGAGGGGCCAGTGAGGTGTGTGACGACTATTACCTGGTACACAGAAACT 234
Db	
	41 LeuSerCysArgThrSerValSer-----SerSerIyrLeuAlaTrpIyIdnIlyPro 59
OY	234 GGCCAGGCTCCCGAGGCTCTCATCATGATGATCAACAGGGCCACTGGATCCGACCC 293
Db	
	60 GlyIleAlaProArgLeuLeuIleTyrlGylAlaSerSerArgAlaThrGlyIleProAsp 79
OY	294 AGTTTCAGTGGCAGTGGGTCTGGGACAGACTCACTCCATCAGACAGACACTGGAGCCC 353
Db	
	80 ArgPheSerGlySerGlySerGlyTyrHisAspHerIleuThrIleSerArgLeuGluPro 99
OY	354 GAAGANGTGGCACTTATTACTGTCCACAATATTATTACATCCCTACACTTTGGCCAG 413
Db	
	100 GluHisSerPheAlaValGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 113



CC	pancreatitis. (I) induces no serious immunorejection due to antigenicity
CC	to human, i.e., human anti-mouse antitigenicity (HMAA) in a host.
CC	AAU74296-AAU74301 represent anti-human ALLm monoclonal antibody amino
XX	acid sequences of the invention.
CC	
SQ	Sequence 236 AA;
Alignment Scores:	
Pred. No.:	3.23e-87
Score:	1098.00
Percent Similarity:	95.34%
Best Local Similarity:	91.53%
Query Match:	62.96%
Dy:	23
Gaps:	2
US-09-806-276A-3 (1-962) x AAU74301 (1-236)	
QY	54 ATGAGACCCCGACCTCAGCTTCCTTCTCCCTGCTACTGGCTGCCAATGCCAGCA 113
Db	1 MetLunhrProLaGlInLeuNeurPheNeuLeuLeuTriPneuroKasphrThcgly 20
QY	114 GAATTGTGTGGACACAGCTCCAGCCACCCTGCTTTTGCTCCAGGGGAAGAGCCACC 173
Db	21 GluLeIleAlLeuThrGlnSerProglyThrLeuSerLeuSerProglyGluAlrAlthr 40
QY	174 CTCTCCGAGGGGCCAGTCAGAGTGTT---AGCAGCTACTTGGCTGGTACCACACAAA 230
Db	41 LeuSerCyArGaLaseRlaserLaserLaserSerSerLeuAlatrPrGlnGlnlys 60
QY	231 CCGGCACAGCTCCCGACGGCTCCTCATATGATGATCATCCACAGGGCCATGCATCCA 290
Db	61 ProGlyGlnAlarProGlyLeuLeuLeuLeuPheGlyAlaLaserSerArgAlatrGlyLeu 80
QY	291 CCCAGCTTCAGTGCAGTGGGTGGGAGACAGACTTCATCTCACATCCAGACAGTGGAG 350
Db	81 AspAqPhneSerGlySerGlySerGlyThraSphenTrhleThrlleSerArgLeuglu 100
QY	351 CCCGAAGATGGGACACTTTATTAAGTGTACGACAATTTTACTACTCCGTAC--ACTTT 407
Db	101 ProGluAspPheAlaValTYrTyrcysGlnGlnPheGlySerSerPrometcysserphe 120
QY	408 GCCCAGGGGACCGAGCTGGAGATCAACACACACTGGGCTGCACATCTGCTTCATCTTC 467
Db	121 GlyGlnGlyThrLysLeuGlnIleLysaGrhThralalaalProSerValPhelePhe 140
QY	468 CGGCATATGATGAGCAGTTGAATCTGGAACTGCTCTGTGTGTGGCCGTGAATAAC 527
Db	141 ProProserSpGlnGlnLeuLysSerGlyThraLaserValValCysLeuLeuasnsn 160
QY	528 TTCATCCGAGAGAGGCCAAAGTAGAGTGAAGGGAGTAACCCCTCCAAATCGGGTAC 587
Db	161 PheTyrrProArgrglunlalylalvalrltrprlysthalaspmsnlaLeuInsertglyasn 180
QY	588 TCCCAGAGAGNGTCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 647
Db	181 SerclnsluservalrheludlnlsasperlysaspherThrtlyserleuSerlerthr 200
QY	648 CTGACGCTGAGCAAAGCAGACTACGAGAAGAACACAAGAGTCTACGCTCGGAAGTACCCAT 707
Db	201 LeuThrleuSerlefyAlaAsprtyGlnLysHisLysValtyrAlacysGluValThrnHis 220
QY	708 CAGGGCTCAGCTGCCCGCTCACAAGAAGCTTCAACAGGGGAGAGTGT 755
Db	221 GlnGlyLeuSerSerProvalThrLysSerPheAsnarGlyGlnCys 236
RESULT 15	
AAR52951	
ID AAR52951 standard; Protein; 234 AA.	
AC AAR52951;	
XX	
DT 27-OCT-1994 (first entry)	



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 18, 2003, 17:19:12 : Search time 10.5579 Seconds  
(without alignments) 5361.850 Million cell updates/sec

Title: US-09-806-276A-3  
Sequence: 1 tcgagccgattcgctcgag.....gaatcttgcacaaaaaaa 962

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cgn2.1/USPTO\_SPOOL/US09806276/runat\_18072003\_160925\_12832/app\_query.fasta\_1.3854  
-DB=Issued\_Patents\_AA -QEXT=fastaa -SOFIX=rai -MINMATCH=0.1 -LOOPL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human4.0.cdi  
-LIST=45 -DOCCALIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MNIEN=0 -MAXLEN=2000000000  
-USER=US09806276.ecgn.1.1.56 -runat\_18072003\_160925\_12832 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGOQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONJLOG  
-DEV\_TIMEOUT=120 -MANY\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1073	61.5	234	4	US-09-049-672A-6
2	1009.5	57.9	235	1	US-08-876-852-153
3	1009.5	57.9	235	1	US-08-876-852-153
4	1009.5	57.9	235	1	US-08-899-575-153
5	1009.5	57.9	235	1	US-08-899-575-153
6	1003	57.5	235	5	PCR-US95-08743-153
7	1001	57.4	240	4	US-08-812-586-16
8	994.5	57.0	241	2	US-09-301-593-36
9	979.5	56.2	214	2	US-07-916-098A-56
10	979.5	56.2	214	2	US-08-480-753-6
11	979.5	56.2	214	3	US-09-041-889-11
12	968.5	55.5	239	3	US-08-837-058-11
					Sequence 6, Appli

13	968	55.5	215	2	US-08-480-753-8	Sequence 8, Appli
14	961	55.1	240	4	US-09-301-593-28	Sequence 28, Appli
15	958	54.9	236	1	US-08-157-101A-5	Sequence 5, Appli
16	955.5	54.8	233	4	US-09-485-737B-69	Sequence 69, Appli
17	951.5	54.6	239	3	US-08-812-586-29	Sequence 29, Appli
18	947.5	54.3	235	4	US-09-171-945-97	Sequence 97, Appli
19	943	54.1	214	4	US-09-679-397-1	Sequence 1, Appli
20	943	54.1	214	4	US-09-680-148-1	Sequence 1, Appli
21	943	54.1	237	2	US-08-463-587A-25	Sequence 25, Appli
22	943	54.1	237	2	US-08-463-587A-25	Sequence 25, Appli
23	943	54.1	237	2	US-08-463-587A-25	Sequence 25, Appli
24	943	54.1	237	2	US-08-463-587A-25	Sequence 25, Appli
25	941	54.0	218	3	US-08-887-352B-13	Sequence 13, Appli
26	941	54.0	218	3	US-08-887-352B-13	Sequence 13, Appli
27	941	54.0	218	3	US-08-887-352B-13	Sequence 13, Appli
28	941	54.0	218	3	US-08-887-352B-13	Sequence 13, Appli
29	941	54.0	218	3	US-08-887-352B-13	Sequence 13, Appli
30	939	53.8	218	4	US-09-296-005-13	Sequence 13, Appli
31	938	53.8	218	4	US-09-296-005-13	Sequence 13, Appli
32	936.5	53.7	233	2	US-07-934-373C-25	Sequence 25, Appli
33	936.5	53.7	233	2	US-07-934-373C-25	Sequence 25, Appli
34	936.5	53.7	233	2	US-07-934-373C-25	Sequence 25, Appli
35	936.5	53.7	233	2	US-07-934-373C-25	Sequence 25, Appli
36	935	53.6	214	4	US-09-247-352-4	Sequence 4, Appli
37	935	53.6	214	4	US-09-247-352-4	Sequence 4, Appli
38	934	53.6	218	2	US-08-887-352B-15	Sequence 15, Appli
39	934	53.6	218	2	US-08-887-352B-15	Sequence 15, Appli
40	934	53.6	218	2	US-08-887-352B-15	Sequence 15, Appli
41	934	53.6	218	2	US-08-887-352B-15	Sequence 15, Appli
42	934	53.6	218	2	US-08-887-352B-15	Sequence 15, Appli
43	934	53.6	218	4	US-09-109-207C-15	Sequence 15, Appli
44	934	53.6	218	4	US-09-109-207C-15	Sequence 15, Appli
45	934	53.6	218	4	US-09-109-207C-15	Sequence 15, Appli

#### ALIGNMENTS

RESULT 1  
US-09-049-672A-6  
Sequence 6, Application US/09049672A  
Patent No. 6135941  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
APPLICANT: Tang, Y. Tom  
APPLICANT: Yue, Henry  
APPLICANT: Au-Young, Janice  
APPLICANT: Corley, Neil C.  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: HUMAN IMMUNE SYSTEM ASSOCIATED PROTEINS  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/049,672A  
FILING DATE: HERWITH  
CLASSIFICATION: 536  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Cerione, Michael C





QY 231 CCTGGCCAGGCTCCAGGCTCCATCATATGATGATCCAGAGGCGCATGCCATCCCA 290  
 Db 61 ProgluylalalProargleuVallleHISGLYValSerAsnArgAlaSerGlylleSer 80  
 QY 291 CCCAGGTGAGTGGAGAGGCTGCTGGGACAGACTTCACCTCACCATCAGCAGACTGGAG 350  
 Db 81 AsparGpHeserGlySerGlySerGlyThrAspHethrLeuThrIleThrArgValGlu 100  
 QY 351 CCCGAGAGATGTCGACTTATATCTGATGATCAATTTTACTACCTCCGTCACCTTTGGC 410  
 Db 101 ProgluAspPheAlaLeuIleTyrCysGlnValIlyrGlyAlaSerSerTyrThrPheGly 120  
 QY 411 CAGGGGACCAAGCTGGAATCAAAAGCAAGCTGCGTCCGACCATCTGCTTCATCTCCG 470  
 Db 121 GlnGlyThrLysLeuGluArgLysArgThrValProAlaProSerValPheIlePhePro 140  
 QY 471 CCATCTGATGAGCACTTGAATCTGGAAGCTGCTCTGTGTGCTCCCTGATTAACCTTC 530  
 Db 141 ProSerAspGlnGlnLeuLysSerGlyThrAlaSerValValCysLeuAsnAspHe 160  
 QY 531 TATCCAGAGAGCCCAAGATACAGTGAAGCTGATACGCCCTCAATCGGGTAACTCC 590  
 Db 161 TyrProArgGluAlaLysValGlnThrLysValAspAsnAlaLeuGlnSerGlyAsnSer 180  
 QY 591 CAGAGAGTGTACAGAGCAGAGCAGAGCAGCAGCCTACAGCCTCAGCAGCAGCCTG 650  
 Db 181 GlnGlnSerValThrGlnGlnAspSerLysAspSerThrTyrSerLeuSerSerThrLeu 200  
 QY 651 ACCTGAGCAAGAGCAGACTACAGAGAAACAAAGTACAGCCTGCGAGAGTCCCACTGAG 710  
 Db 201 ThrLeuSerLysAlaAspTyrGlnLysHisLysValIlyrAlaCysGlnValThrHisGln 220  
 QY 711 GGCTGAGCTGCGCCCTGCACAAAGAGCTTCAACAGGGGAGAGTGT 755  
 Db 221 GlyLeuSerSerProValThrLysSerPheAsnArgGlyGlnCys 235

RESULT 3  
 US-08-899-575-153  
 ; Sequence 153, Application US/08899575  
 ; Patent No. 5770440  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Burton, Dennis R  
 ; APPLICANT: Barbas, Carlos F  
 ; APPLICANT: Lerner, Richard A  
 ; TITLE OF INVENTION: HUMAN NEUTRALIZING MONOCLONAL ANTIBODIES  
 ; TITLE OF INVENTION: TO HUMAN IMMUNODEFICIENCY VIRUS  
 ; NUMBER OF SEQUENCES: 170  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: The Scripps Research Institute, Office of  
 ; ADDRESSEE: Patent Counsel  
 ; STREET: 10666 No. 5770440th Torrey Pines Road, Suite 220,  
 ; STREET: Mail Drop 7PC8  
 ; CITY: La Jolla  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 92037  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/899, 575  
 ; FILING DATE: 24-JUL-1997  
 ; CLASSIFICATION: 435  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/276, 852  
 ; FILING DATE: 18-JUL-1994  
 ; APPLICATION NUMBER: US 08/178, 302  
 ; FILING DATE: 30-SEP-1993  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/954, 148

; FILING DATE: 30-SEP-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Filling, Thomas  
 ; REGISTRATION NUMBER: 34,163  
 ; REFERENCE/DOCKET NUMBER: SCRI452P  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 619-554-2937  
 ; TELEFAX: 619-554-6312  
 ; INFORMATION FOR SEQ ID NO: 153:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 235 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-899-575-153

Alignment Scores:  
 Pred. No.: 7,69e-91 Length: 235  
 Score: 1009.50 Matches: 198  
 Percent Similarity: 88.94% Conservative: 11  
 Best Local Similarity: 84.26% Mismatches: 25  
 Query Match: 57.88% Indels: 1  
 DB: 1 Gaps: 1

US-09-806-276A-3 (1-962) x US-08-899-575-153 (1-235)

QY 54 ATGGAAGCCCGAGCTGCTGCTTCTCTCTCTGCTGCTGCTCCAGATACCAACCGGA 113  
 Db 1 MetGlyValProThrGlnValLeuGlyLeuLeuLeuThrPheThrAspAlaArgCys 20  
 QY 114 GAAATTTGTTGACAGCTGCTCCAGCCACCCCTGCTTTGTCTCCAGAGGAAAGACCAAC 173  
 Db 21 GluIleValLeuThrGlnInsProGlyThrLeuSerLeuSerProGlyGlnArgAlaThr 40  
 QY 174 CTCTCTGCGAGGCGCAGCTCAGAGTGT--AGCAGCTACTTACGCTGAGTACCAAGAA 230  
 Db 41 PheSerCysArgSerSerHisSerIleArgSerArgArgValAlaThrPyrGlnHisLys 60  
 QY 231 CCGGCGAGGCTCCAGGCTCCCTCATCTATGATGATTCACCAAGGCGCACTGGCATCCCA 290  
 Db 61 ProGlyGlnAlaProArgLeuValIleHisGlyValSerAsnArgAlaSerGlylleSer 80  
 QY 291 CCCAGGTGAGTGGAGAGGCTGCTGGGACAGACTTCACCTCAGCATCAGAGAGTGGAG 350  
 Db 81 AsparGpHeserGlySerGlySerGlyThrAspHethrLeuThrIleThrArgValGlu 100  
 QY 351 CCCGAGAGTGGCAGCTTATATCTGATGATCAATTTTACTACCTCCGTCACCTTTGGC 410  
 Db 101 ProgluAspPheAlaLeuIleTyrCysGlnValIlyrGlyAlaSerSerTyrThrPheGly 120  
 QY 411 CAGGGGACCAAGCTGGAATCAAAAGCAAGCTGCGTCCGACCATCTGCTTCATCTCCG 470  
 Db 121 GlnGlyThrLysLeuGlnArgLysArgThrValProAlaProSerValPheIlePhePro 140  
 QY 471 CCATCTGATGAGCACTTGAATCTGGAAGCTGCTCTGTGTGCTCCCTGATTAACCTTC 530  
 Db 141 ProSerAspGlnGlnLeuLysSerGlyThrAlaSerValValCysLeuAsnAspHe 160  
 QY 531 TATCCAGAGAGCCCAAGATACAGTGAAGTGAAGTACGCCCTCAATCGGGTAACTCC 590  
 Db 161 TyrProArgGluAlaLysValGlnThrLysValAspAsnAlaLeuGlnSerGlyAsnSer 180  
 QY 591 CAGAGAGTGTACAGAGCAGAGCAGAGCAGCAGCCTACAGCCTCAGCAGCAGCCTG 650  
 Db 181 GlnGlnSerValThrGlnGlnAspSerLysAspSerThrTyrSerLeuSerSerThrLeu 200  
 QY 651 ACCTGAGCAAGAGCAGACTACAGAGAAACAAAGTACAGCCTGCGAGAGTCCCACTGAG 710  
 Db 201 ThrLeuSerLysAlaAspTyrGlnLysHisLysValIlyrAlaCysGlnValThrHisGln 220  
 QY 711 GGCTGAGCTGCGCCCTGCACAAAGAGCTTCAACAGGGGAGAGTGT 755  
 Db 221 GlyLeuSerSerProValThrLysSerPheAsnArgGlyGlnCys 235



RESULT 6  
 US-08-812-586-16  
 : Sequence 16, Application US/08812586  
 : Patent No. 6048704  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Martin David Tilson  
 : TITLE OF INVENTION: PURIFIED AND RECOMBINANT ANTIGENIC  
 : TITLE OF INVENTION: PROTEINS ASSOCIATED WITH ABDOMINAL AORTIC ANEURYSM (AAA)  
 : TITLE OF INVENTION: DISPEASE, AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF  
 : NUMBER OF SEQUENCES: 61  
 :  
 : CORRESPONDENCE ADDRESS:  
 :  
 : ADDRESSEE: Cooper & Dunham LLP  
 : STREET: 1185 Avenue of the Americas  
 : CITY: New York  
 : STATE: New York  
 : COUNTRY: U.S.A.  
 : ZIP: 10036  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.30  
 :  
 : CURRENT APPLICATION DATA:

QY	54	ATGGAAACCCACAGTACGGCTCTTCCTTCCTGCTGACACTGGCTCCGAGATACCAACCGGA	113
Db	3	MetwrgalAprohlaenleuLeuEnglyLeuLeuLeuTrpLeuProglyAlaArgCys	22
QY	114	GAATTTGTGTGACAACAAGTCTCCAGCCACCTGTCTTGTCTCCAGGGGAAAGAGCCAC	173
Db	23	AlaIleArgIleAlaGlnSerProSerSerLeuSerAlaSerThrGlyAspArgValThr	42
QY	174	CTCTCCAGCGGGCCAGTCAGAGTGTATTACACTCTTACGCTGGTATACCAACAACCT	233
Db	43	IleThrCysArgAlaSerGlnGlyIleSerAsnTrpLeuAlaIleTPYrGlnGlnLysPro	62
QY	234	GGCCAGGCTCCGAGGCTCCTCATCTATGATGATCAACAGGGCCACTGGCATCCACCC	293
Db	63	GlyGlnIleArgProArgLeuLeuIleTyrAspAlaSerSerArgAlaThrGlyIleProAsp	82
QY	294	AGGTTACAGTGGCAGTGGTGTGGGACAGACTCACCTGCACATCAGACAGCTGGAGCC	353
Db	83	ArgPheSerGlySerGlySerGlyThrAspThrLeuThrIleSerArgLeuGlnPro	102
QY	354	GAAGATGTGGCACTTATTACCTGACGACAATTTTACTACTCCGTCACCTTTGGCCAG	413
Db	103	GlnAspPheAlaValIleTyrTrpGlnGlnGlnTyrGlySerProLeuThrPheGlyGly	122
QY	414	GGGACACAGGCTGGAGATCAACGAACCTGGGTGGCACCANTGTCTTCATCTCCGCCA	473
Db	123	GlyThrIleValGlnIleLeuSarphrValAlaAlaProSerValPheIlePheProPro	142
QY	474	TTCGATGAGCAGTGGAAATCTGAGACTCCCTCTGTGTGTGCTTCGCTGAATACCTCTAT	533
Db	143	SerAspGlnGlnLeuLysSerGlyThrAlaSerValAlaGlyLeuLeuAsnAsnPheTyr	162
QY	534	CCCAGAGAGCCAAAGTACAGTGAAGGTGATTAAGCGCTCCACANTGGGTAACTCCCA	593
Db	163	ProArgGlnAlaIleValGlnTrpPylSerValAspAsnAlaLeuGlnSerGlyAsnSerGln	182
QY	594	GAGAGTGTACAGAGAGAGACAGCAAGAGCAAGCAACCCACACACACCTGAG	653
Db	183	GlnSerValThrGlnGlnAspSerLysAspSerThrTyrSerLeuSerSerThrLeuThr	202
QY	654	CTGAGCAAAAGCAGACTACGAGAAACACAAAGTCTGAGCCTCGGAAGTACACCATCAGGC	713
Db	203	LeuSerLysAlaAspTyrGlnLysHisLysValTyrAlaGlyAlaGlnValThrHisGlnGly	222







## Alignment Scores:

Pred. No.: 6.45e-88 Length: 214  
 Score: 979.50 Matches: 190  
 Percent Similarity: 93.87% Conservative: 9  
 Best Local Similarity: 89.62% Mismatches: 12  
 Query Match: 56.16% Indels: 1  
 DB: 3 Gaps: 1

US-09-806-276a-3 (1-962) x US-08-837-058-11 (1-214)

```

OY 123 TTGACAGAGTGTCCAGCCACCCCTCTTGTGCTCCAGGGAGAAAGCCACCCCTCTCTGC 182
    |||||
Db 3 LeuThrgInSerProGlyThrLeuSerLeuPheProGlyGluArgAlaThrLeuSerCys 22
    |||||
OY 183 AGGCGCAGTCAAGTGTAGC---AGCTACTTACCTGTGACCAAGCAAGAACCTGGGCGAG 239
    |||||
Db 23 ArgAlaSerGlnArgIleSerThrSerPheGlnAlaTrpGlnGlnLysProGlyGln 42
    |||||
OY 240 GCTCCAGAGCTCCCTCATCTATGATGATCCAAACAGGGCCACTGGCATCCACCCAGGTTG 239
    |||||
Db 43 SerProArgLeuLeuIlePheAspAlaSerThrArgAlaProGlyIleProAspArgPhe 62
    |||||
OY 300 AGTGCAGTGGGTGTGGGACAGACTTCATCTCACCATCAGCAGACTGGAGCCGGAAGAT 339
    |||||
Db 63 SerAlaSerTrpSerGlyThrAspPheThrLeuThrIleSerArgLeuGluProGlyLys 82
    |||||
OY 360 GTGGCACTTATTAATCTGACCAATATTTACTACCTCCATCTTGGCCAGGGGAGC 419
    |||||
Db 83 PheAlaValIleTyrCysGlnHisTyrGlySerProTrpTrpPheGlyGlnGlyThr 102
    |||||
OY 420 AGGCTGGAGATGCAACCAAGTGGCTGGCCACCATCTGTCTTCACTCTCCCGCATGTGAT 479
    |||||
Db 103 LysAlaIleValIleLysAlaGlnValAlaAlaProSerValPheIlePheProProSerAsp 122
    |||||
OY 480 GAGCAGTTGAATCTGGAACCTGCTGTGTGCTGCTGCTGAATTAATCTTATCCACAG 539
    |||||
Db 123 GlnGlnIleLysSerGlyThrAlaSerValValCysLeuLeuAsnAsnPheTyrProArg 142
    |||||
OY 540 GAGGCCAAATATACAGTGGAGGTGATACGCCCTCCATCGGGTAACTCCACAGAGAT 599
    |||||
Db 143 GluAlaIleValIleTrpLysValAspAsnAlaLeuGlnSerGlyLysSerGlnGlnSer 162
    |||||
OY 600 GTCCAGAGCAGGACGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 659.
    |||||
Db 163 ValThrGlnGlnAspSerLysAspSerThrTyrSerLeuSerSerThrLeuThrLeuSer 182
    |||||
OY 660 AAAGCAGACTACAGAGAACAACAAGTCTACGCTCGCAAGTCAACCATCAGGCGCTGAGC 719
    |||||
Db 183 LysAlaAspTyrGlnLysHisLysValTyrAlaCysGluValThrHisGlnGlyLeuSer 202
    |||||
OY 720 TCGCCCGTCAACAAGCTTCAACAGGAGAGGTGT 755
    |||||
Db 203 SerProValThrLysSerPheAsnArgGlyGlnCys 214
    |||||

```

## RESULT 12

US-08-487-550-6  
 ; Sequence 6, Application US/08487550  
 ; Patent No. 6113898

## GENERAL INFORMATION:

APPLICANT: Anderson, Darrell R.  
 TITLE OF INVENTION: "MONKEY MONOCLONAL ANTIBODIES SPECIFIC  
 TO HUMAN B7.1 AND/OR B7.2 PRIMATIZED FORMS THEREOF,  
 TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS CONTAINING, AND USE THEREOF AS  
 IMMUNOSUPPRESSANTS"  
 NUMBER OF SEQUENCES: 12

## CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
 STREET: 699 Prince Street  
 CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22314

COMPUTER READABLE FORM:

## MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487,550

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030

REFERENCE/DOCKET NUMBER: 012712-131

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6620

TELEFAX: 703-836-2021

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 239 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-487-550-6

## Alignment Scores:

Pred. No.: 8.09e-87 Length: 239  
 Score: 968.50 Matches: 187  
 Percent Similarity: 87.87% Conservative: 23  
 Best Local Similarity: 78.24% Mismatches: 24  
 Query Match: 55.53% Indels: 5  
 DB: 3 Gaps: 1

US-09-806-276a-3 (1-962) x US-08-487-550-6 (1-239)

```

OY 54 ATGGAAGCCCGCAGCTTCTTCTTCTCTGCTGCTGCTCCAGATACCAACCGGA 113
    |||||
Db 1 MetSerLeuProAlaGlnLeuLeuGlnLeuLeuLeuLeuValProGlySerSerGly 20
    |||||
OY 114 GAAATGTGTTGACACAGCTCTCCAGCCACCGCTGTCTTGTCCAGGGGAAAGCCAGC 173
    |||||
Db 21 GluValAlaMetThrGlnSerProLeuSerLeuProIleThrProGlyGlnProAlaSer 40
    |||||
OY 174 CTCTCTCGAGGGCCAGTCAAGTGT-----AGCAGCTACTTACCTGG 218
    |||||
Db 41 IleSerCysArgSerSerGlnSerLeuLysHisSerAsnGlyAspThrPheLeuSerTrp 60
    |||||
OY 219 TACCAACAAGAAACCTGGCCAGGCTCCAGGCTCTCATCTTATGATCATCCACAGGCC 278
    |||||
Db 61 TyrGlnGlnLysProGlyGlnProArgLeuLeuIleTyrLysValSerAsnArgAsp 80
    |||||
OY 279 ACTGGCATCCACCCAGGTTGATGGCAGTGGGTGGTGGAGACACTTCACCTCAGCATC 338
    |||||
Db 81 SerGlyValProAspArgPheSerGlySerIleAlaGlyThrAspPheThrLeuLysIle 100
    |||||
OY 339 AGCAGACTGAGACCCGGAAGATGTGGCACTTATTACTGTCAACAATATTTACTACTCCG 398
    |||||
Db 101 SerAlaValGluAlaGluAspValGlyValTyrPheCysGlyGlnGlyThrArgThrPro 120
    |||||
OY 399 TACACTTTGGCCAGGGGAGCCAGGCTGAGATTCAAAGAACTGTGGCTGACCATCTGTC 458
    |||||
Db 121 ProThrPheGlyGlyGlyThrLysValGlnIleLysArgThrValAlaAlaProSerVal 140
    |||||
OY 459 TTCACTCTCCGCGCATCTGATGACAGTGAATCTGAATCTGCACTGCTGTGGTGGCTG 518
    |||||
Db 141 PheIlePheProProSerAspGlnGlnLeuLysSerGlyThrAlaSerValValCysLeu 160
    |||||
OY 519 CTGAATTAATCTTATCCAGAGAGCCCAAGTACAGTGAAGGTGATAAGCCCTCCAA 578
    |||||
Db 161 LeuAsnAsnPheTyrProArgGlnAlaLysValGlnTrpLysValAspAsnAlaLeuGln 180
    |||||
OY 579 TCGGGTACTCCAGAGAGTGTACAGAGAGAGAGCAAGAGACACACCTACAGCTC 638
    |||||
Db 181 SerGlyAsnSerGlnGlnSerValThrGlnLysPheLysAspSerTrpTyrSerLeu 200
    |||||

```





```

QY 114 GAAATGTGGTGAACAGAGTGTCCAGGCAACCTGTCTTTCTCCAGGGGAAAGACCCACC 173
Db 21 AsplleValMetSerGlnSerProSerSerLeuAlaValSerValGlyGlnLysValThr 40
QY 174 CTCTCTCGACGAGGGCCACTGACAGTGT-----AGCAGCTACTTGAACC 215
Db 41 MetSerGlySerSerGlnSerGlnSerLeuLeuTyrSerArgAsnGlnLysAsnTyrLeuAla 60
QY 216 TGTGTACCAACAGAAACTGTGGCCAGGCTCCAGGCTCTCATATTATGATGATTCACACAGG 275
Db 61 TTPheGlnGlnLysProGlyGlnSerProLysLeuLeuIlePheTrpAlaSerThrArg 80
QY 276 GCACCTGCGAATCCCAACCCAGTTCAGTGGAGAGGGGTGGGAGCAGACTTCACCTCAC 335
Db 81 GlnSerGlyAlaProAspArgPheThrGlySerGlyPheGlyThrAspPheAsnLeuThr 100
QY 336 ATCAGCAGACTGGAGCCCGAAGATGTGGACATTATTACTGTACAGCAATATTTTACTACT 395
Db 101 IleSerSerValGlnAlaGlnAspLeuAlaValTyrAspCysGlnGlnIntYrPheSerTyr 120
QY 396 CCGTACACTTTTGGCCAGGGGAGCAGGCTGGAGATTCAAACAGACTGTGGCTGCACATCT 455
Db 121 ProLeuThrPheGlyAlaGlyThrLysLeuGlnLeuLysValGlyThrValAlaAlaProSer 140
QY 456 GTCTTCATCTTCCCGCATCTGATGAGCAGTGAATCTGGAACCTGCACTGCTGTGTGTGC 515
Db 141 ValPheIlePheProProSerAspGlnGlnLeuLysSerGlyThrAlaSerValValCys 160
QY 516 CTGTGTAATTAATCTTATCTCCAGAGAGGCCAAAGTACAGTGAAGTGAATGATACGCCCTC 575
Db 161 LeuLeuAsnAsnPheTyrProArgGlnAlaLysValGlnTrpLysValAspAsnAlaLeu 180
QY 576 CAATCGGCTAAGTCCAGAGAGATGTCACAGAGCAGAGCAGAGCAGAGCAGACCTACAGC 635
Db 181 GlnSerGlyAsnSerGlnGlnSerValThrGlnGlnAspSerLysAspSerThrTyrSer 200
QY 636 CTCAGCAGACCCCTGACGCTGAGCAAGCAGACACTACGAAACACAAAGTCTACGCCCTGC 695
Db 201 LeuSerSerThrLeuThrLeuSerThrLysAlaAspTyrGlyLysThrLysLysValTyrAlaCys 220
QY 696 GAAGTACCCCATGAGGCGCTGAGCTGAGCTGCCCGGCACAAAGACTTCACAGGGGAGAGTGT 755
Db 221 GlnValThrHisGlnIleLysLeuSerProValThrLysSerPheAsnArgGlyGlyLys 240

RESULT 15
US-08-157-101A-5
: Sequence 5, Application US/08157101A
: Patent No. 5808032
: GENERAL INFORMATION:
: APPLICANT: KURIHARA, TATSUYA
: APPLICANT: MATSUKURA, SHIGEKAZU
: APPLICANT: TSURUOKA, NOBUO
: APPLICANT: ARIWA, KENJI
: APPLICANT: NISHIHARA, TATSURO
: TITLE OF INVENTION: ANTI-HBS ANTIBODY GENES AND EXPRESSION
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PILLSBURY, MADISON & SUTRO
: STREET: 1100 NEW YORK AVENUE, N.W.
: CITY: WASHINGTON
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/157,101A
: FILING DATE: 05-APR-1994
: CLASSIFICATION: 530

```

[illegible]

Fri Jul 25 09:59:14 2003

us-09-806-276a-3.ra1

Page 12

Db 223 LeuSerSerProValThrLySerPheAsnArgGlyGluCys 236

Search completed: July 18, 2003, 17:39:49  
Job time : 16.5579 secs

1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 18, 2003, 17:24:58 ; Search time 35.9242 Seconds

(without alignments)  
6360.447 Million cell updates/sec

Title: US-09-806-276A-3

Perfect score: 17/44  
Sequence: 1 tcgagccgattcgctcgag.....gaattcttgcacaaaaaaaaa 962

Scoring table:

BLOSUM62	
Xgapop 10.0 , Ygapext 0.5	
Egapop 10.0 , Ygapext 0.5	
Egapop 6.0 , Ygapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 903798

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p model -DEV=xlp  
-O=/cgna\_1/USPTO.spool/US09806276/rnat.18072003.160926.12862/app.query.fasta.1.3854  
-DB=Published.Applications\_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62  
-TRANS=human40.cdi -LIST=45 -DOCCALIGN=200 -THR.SCORE=pct -THR.MAX=100  
-THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09806276.ecgn.1.1.42-rnat.18072003.160926.12862  
-NCPU=6 -ICPU=3 -NO.MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100  
-LONLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Published.Applications\_AA.\*  
1: /cgna2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
2: /cgna2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgna2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgna2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgna2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
6: /cgna2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgna2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
8: /cgna2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
9: /cgna2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
10: /cgna2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
11: /cgna2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
12: /cgna2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgna2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
14: /cgna2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
15: /cgna2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
16: /cgna2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
17: /cgna2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1133.5	65.0	235	US-10-153-382-7	Sequence 7, Appl1

2	1127	64.6	234	15	US-10-153-382-15	Sequence 15, Appl
3	1114.5	63.9	233	15	US-10-153-382-11	Sequence 11, Appl
4	1108	63.5	236	11	US-09-859-053-34	Sequence 34, Appl
5	1098	63.0	236	11	US-09-859-053-38	Sequence 38, Appl
6	1053	60.4	238	15	US-10-216-484-107	Sequence 107, Appl
7	1047	60.0	238	15	US-10-216-484-107	Sequence 107, Appl
8	1033	59.2	224	12	US-09-453-234-82	Sequence 82, Appl
9	1033	59.2	224	12	US-09-453-234-88	Sequence 88, Appl
10	1033	59.2	224	12	US-09-453-234-90	Sequence 90, Appl
11	1029	59.0	224	12	US-09-453-234-36	Sequence 36, Appl
12	1029	59.0	238	15	US-10-216-484-52	Sequence 52, Appl
13	1029	59.0	238	15	US-10-216-484-109	Sequence 109, Appl
14	1026	58.8	226	12	US-09-453-234-86	Sequence 86, Appl
15	1023	58.7	226	12	US-09-453-234-50	Sequence 50, Appl
16	1023	58.7	226	12	US-09-453-234-80	Sequence 80, Appl
17	1022.5	58.6	215	12	US-09-453-234-84	Sequence 84, Appl
18	1021	58.5	224	12	US-09-453-234-84	Sequence 84, Appl
19	1021	58.5	238	15	US-10-216-484-54	Sequence 54, Appl
20	1019	58.4	224	12	US-09-453-234-46	Sequence 46, Appl
21	1014	58.1	224	12	US-09-453-234-74	Sequence 74, Appl
22	1012	58.0	226	12	US-09-453-234-38	Sequence 38, Appl
23	1008	57.8	226	12	US-09-453-234-42	Sequence 42, Appl
24	1003.5	57.5	224	12	US-10-221-945-1	Sequence 1, Appl
25	1003	57.5	224	12	US-09-453-234-44	Sequence 44, Appl
26	1003	57.5	224	12	US-09-453-234-78	Sequence 78, Appl
27	1003	57.5	238	15	US-10-216-484-129	Sequence 129, Appl
28	1002	57.5	224	12	US-09-453-234-52	Sequence 52, Appl
29	1001	57.4	226	12	US-09-453-234-72	Sequence 72, Appl
30	1000	57.3	238	15	US-10-216-484-131	Sequence 131, Appl
31	999	57.3	238	15	US-10-216-484-127	Sequence 127, Appl
32	999	57.3	240	10	US-09-799-514-8	Sequence 8, Appl
33	998	57.2	215	12	US-09-791-153A-47	Sequence 47, Appl
34	996.5	57.1	234	15	US-10-026-925-55	Sequence 55, Appl
35	994	57.0	226	11	US-09-859-053-30	Sequence 30, Appl
36	993	56.9	224	12	US-09-453-234-40	Sequence 40, Appl
37	992	56.9	224	12	US-09-453-234-76	Sequence 76, Appl
38	991.5	56.9	239	11	US-09-249-011A-72	Sequence 22, Appl
39	991.5	56.9	239	12	US-09-992-600A-8	Sequence 8, Appl
40	991.5	56.9	239	12	US-09-924-340-8	Sequence 8, Appl
41	991.5	56.9	239	15	US-10-000-989-8	Sequence 8, Appl
42	991.5	56.8	239	15	US-10-000-989-8	Sequence 8, Appl
43	991.5	56.8	238	15	US-10-171-452A-38	Sequence 38, Appl
44	991	56.8	238	15	US-10-171-452A-56	Sequence 56, Appl
45	989	56.7	234	10	US-09-740-002-24	Sequence 24, Appl

## ALIGNMENTS

RESULT 1  
US-10-153-382-7  
Sequence 7, Application US/10153382  
Publication No. US20030086930A1  
GENERAL INFORMATION:  
APPLICANT: PFIZER PRODUCTS INC.  
TITLE OF INVENTION: USES OF ANTI-CTLA-4 ANTIBODIES  
FILE REFERENCE: PC23019A  
CURRENT APPLICATION NUMBER: US/10/153,382  
CURRENT FILING DATE: 2002-05-22  
PRIOR APPLICATION NUMBER: 60/293042  
PRIOR FILING DATE: 2001-05-23  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 235  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-153-382-7

Alignment Scores:  
Pred. No.: 2.4e-79  
Score: 1133.50  
Percent Similarity: 97.02%  
Best Local Similarity: 93.19%  
Length: 235  
Matches: 219  
Conservative: 9  
Mismatch: 6







```

QY 402 ACTTTGGCCAGGGGAGCGCTGAGATCAAGAACTGTGGCTGCACCCATCTGCTTC 461
    |||||||
Db 121 Thrheglyngllythrlylsleugluilelysargthrvalalaalaproservalphe 140
QY 462 ATCTCCCGCCATCTGATGAGCACTGTAATCTGGAACCTCCTCTGTGTGCTGCTG 521
    |||||||
Db 141 lleheproproseraspjnglinsleulysserglythralservalvalcysleu 160
QY 522 AATACCTTATATCCAGAGAGGCCAAGTACAGTGAAGGTAGTACGCCCTCCATCG 581
    |||||||
Db 161 Asnasnphetryproalrgglualalysvalglntrplysvalaspsnalaleuglnser 180
QY 582 GGTAACTCCAGAGAGCTGTACAGAGACAGACAGCAAGACAGCAACCTACAGCTCAG 641
    |||||||
Db 181 Glysenserlnglnservalthrclnglnasperlysasperthrlyrserleuser 200
QY 642 AGCACCTGACGCTGAGCAAGCAAGCACTACAGAAACAAAGCTACGCTCGCAATCG 701
    |||||||
Db 201 Serthrleuthrleuserlysalasptyrglulshlyslvaltyralacysgluval 220
QY 702 ACCCATGAGGCGCTGAGCTGCGCCGTCAACAAGAGCTTCAACAGGGAGAGTGT 755
    |||||||
Db 221 Thrhislngllyleuserserprovalthrlysserpheasnarglyglucys 238

RESULT 7
US-10-216-484-50
: Sequence 50, Application US/10216484
: Publication No. US20030103976A1
: GENERAL INFORMATION:
: APPLICANT: Serizawa, No. US20030103976A1ufusa
: APPLICANT: Haruyama, Hideyuki
: APPLICANT: Nakahara, Kaori
: APPLICANT: Tamaki, Ikuro
: APPLICANT: Takahashi, Tohru
: TITLE OF INVENTION: Anti-Fas Antibodies
: FILE REFERENCE: 980126CJP/HG
: CURRENT APPLICATION NUMBER: US/10/216,484
: PRIOR FILING DATE: 2002-08-09
: PRIOR APPLICATION NUMBER: US/09/499,662
: PRIOR FILING DATE: 2000-02-09
: PRIOR APPLICATION NUMBER: US 09/053,583
: PRIOR FILING DATE: 1998-04-01
: NUMBER OF SEQ ID NOS: 165
: SEQ ID NO 50
: LENGTH: 238
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Designed light
US-10-216-484-50

Alignment Scores:
Pred. No.: 1,12e-72 Length: 238
Score: 1047.00 Matches: 206
Percent Similarity: 90.34% Conservative: 9
Best Local Similarity: 86.55% Mismatches: 19
Query Match: 60.03% Indels: 4
DB: 15 Gaps: 1

US-09-806-276a-3 (1-962) x US-10-216-484-50 (1-238)
QY 54 ATGGAAGCCCGACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
    |||||||
Db 1 Metglthrashprllleuleutrpvalleuleutripvalprogllyserthrly 20
QY 114 GAATATGTGTGACACAGTCTCAGCCACCCCTCTCTTGTCTCCAGGGAAGAGCCACC 173
    |||||||
Db 21 Aspilevalleuthrleuserprogllythrleuserleuserprogllyluatgalatthr 40
QY 174 CTCTCCTGAGGCGCGAGTCAAGAGTGT-----AGACGCTACTAGCTGTGATC 221
    |||||||

```

```

Db 41 Leusercyslysalaserclnservalasplyraasglyaspsertyrmetasnrptlyr 60
QY 222 CAACAGAAACCTGGCCAGGCTCCAGCGCTCCCTCATCTATGATGATCCAAAGGCCACT 281
    |||||||
Db 61 Glnghlnysprogllylnalaproargleuleutllyrvalalaaserasmeuglnser 80
QY 282 GCCATCCCGCCAGGCTGAGTGGAGTGGCTGGAGACAGACTTCACCTCCACATCAGC 341
    |||||||
Db 81 Glylleproaspargpheserlyserglyserglythrlnasphethrleuthrleuser 100
QY 342 AGACTGAGCCCGAAGATGTCGCACTTATTAATCTGACCAATATTTACTACTCCGTAC 401
    |||||||
Db 101 Argleuglnproalasphehalavaltyrtyrcysglnlnserasngl uasproarg 120
QY 402 ACTTTGGCCAGGGGAGCGCTGAGATCAAGAACTGTGGCTGCACCCATCTGCTTC 461
    |||||||
Db 121 Thrheglyngllythrargleugluilelysargthrvalalaalaproservalphe 140
QY 462 ATCTCCCGCCATCTGATGAGCACTGTAATCTGGAACCTCCTCTGTGTGCTGCTG 521
    |||||||
Db 141 lleheproproseraspjnglinsleulysserglythralservalvalcysleu 160
QY 522 AATACCTTATATCCAGAGAGGCCAAGTACAGTGAAGGTAGATTAACGCCCTCCATCG 581
    |||||||
Db 161 Asnasnphetryproalrgglualalysvalglntrplysvalaspsnalaleuglnser 180
QY 582 GGTAACTCCAGAGAGCTGTACAGAGACAGACAGCAAGACAGCAACCTACAGCTCAG 641
    |||||||
Db 181 Glysenserlnglnservalthrclnglnasperlysasperthrlyrserleuser 200
QY 642 AGCACCTGACGCTGAGCAAGCAAGCACTACAGAAACAAAGCTACGCTCGCAATCG 701
    |||||||
Db 201 Serthrleuthrleuserlysalasptyrglulshlyslvaltyralacysgluval 220
QY 702 ACCCATGAGGCGCTGAGCTGCGCCGTCAACAAGAGCTTCAACAGGGAGAGTGT 755
    |||||||
Db 221 Thrhislngllyleuserserprovalthrlysserpheasnarglyglucys 238

RESULT 8
US-09-453-234-82
: Sequence 82, Application US/09453234
: Publication No. US20030091995A1
: GENERAL INFORMATION:
: APPLICANT: Buechler, Joe
: APPLICANT: Walkers, Gunars
: APPLICANT: Gray, Jeff
: APPLICANT: Lomborg, Nils
: APPLICANT: Biosite Diagnostics, Inc.
: APPLICANT: GenPharm International
: TITLE OF INVENTION: Human Antibodies
: FILE REFERENCE: 020015-000110US
: CURRENT APPLICATION NUMBER: US/09/453,234
: PRIOR FILING DATE: 1999-12-01
: PRIOR APPLICATION NUMBER: US 60/157,415
: PRIOR FILING DATE: 1999-10-02
: NUMBER OF SEQ ID NOS: 112
: SOFTWARE: Patentln Ver. 2.1
: SEQ ID NO 82
: LENGTH: 224
: TYPE: PRT
: ORGANISM: Homo sapiens
: OTHER INFORMATION: M2-31L
US-09-453-234-82

Alignment Scores:
Pred. No.: 1.33e-71 Length: 224
Score: 1033.00 Matches: 202
Percent Similarity: 96.24% Conservative: 3
Best Local Similarity: 94.84% Mismatches: 8
Query Match: 59.23% Indels: 0
DB: 12 Gaps: 0

US-09-806-276a-3 (1-962) x US-09-453-234-82 (1-224)

```





TYPE: PRT  
 ORGANISM: Homo sapiens  
 OTHER INFORMATION: M2-35L  
 US-09-453-234-90

## Alignment Scores:

Pred. No.: 1,33e-71 Length: 224  
 Score: 1033.00 Matches: 202  
 Percent Similarity: 96.24% Conservative: 3  
 Best Local Similarity: 94.84% Mismatches: 8  
 Query Match: 59.23% Indels: 0  
 DB: 12 Gaps: 0

US-09-806-276a-3 (1-962) x US-09-453-234-90 (1-224)

```

QY 114 GAAATGTTGACACAGTCCTCCAGCCCTGCTTTGTCCTCCAGGGAAAGACCACC 173
   |||||
Db 1 GluileValLeuThrGlnSerProAlaThrLeuSerLeuSerProGlyLysArgAlaThr 20
QY 174 CTCCTCTGACAGGCGCAGTCAGAGTGTAGCAGCTTACCTGTGACCAAGAAACCT 233
   |||||
Db 21 LeuSerCysArgAlaSerGlnSerValSerSerTyrLeuAlaIlePtyrGlnGlnLysPro 40
QY 234 GGCAGGCGTCCAGGCTCTCTCATATGATGCATCCAAAGGCGCAGTGGCATCCACC 293
   |||||
Db 41 GlyLysAlaProArgLeuLeuIleTyrAspAlaSerAsnArgAlaThrGlyIleProAla 60
QY 294 AGGTTACAGTGCAGTGGTCTGGAGACAGCTTACCTCACCACATCAGACAGACTGGAGCC 353
   |||||
Db 61 ArgPheSerGlySerGlySerGlyThrAspPheThrLeuThrIleSerSerLeuGlnPro 80
QY 354 GAAAGTGTGACACTTATTACTGTACAGAAATATTTTACTACTCGTACACTTTGGCCAG 413
   |||||
Db 81 GluAspPheAlaValIleTyrCysGlnGlnArgThrAsnTrpProArgThrPheGlyGln 100
QY 414 GGGACAGGCTGGAGATCAAGAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 473
   |||||
Db 101 GlyThrLysValGluIleLysArgThrValAlaAlaProSerValPheIlePheProPro 120
QY 474 TCTGATGACAGCTGAAATCTGGAATGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTAT 533
   |||||
Db 121 SerAspGlnGlnLeuLysSerGlyThrAlaSerValIleCysLeuLeuAsnAsnPheTyr 140
QY 534 CCCAGAGAGGCGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 593
   |||||
Db 141 ProArgLysAlaLysValGlnTrpLysValAlaAspAsnAlaLeuGlnSerGlyAsnSerGln 160
QY 594 GAGAGTGTACAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAG 653
   |||||
Db 161 GluSerValThrGlnGlnAspSerLysAspSerThrTyrSerLeuSerSerThrLeuThr 180
QY 654 CTGAGCAAGCAGACTACGAGAAACACAAAGTCTACCGCTGCGAAGTCAACCCATCAGGCG 713
   |||||
Db 181 LeuSerLysAlaAspTyrGlnLysLysIleValTyrAlaCysGlnValThrHisGlnGly 200
QY 714 CTGAGCTGCGCGCTCAAAAGAGCTTCAACAGGGGAGAG 752
   |||||
Db 201 LeuSerSerProValThrLysSerPheAsnArgGlyGln 213

```

## RESULT 11

US-09-453-234-36  
 Sequence 36, Application US/09453234  
 Publication No. US20030091995A1  
 GENERAL INFORMATION:  
 APPLICANT: Buechler, Joe  
 APPLICANT: Valkirs, Gunars  
 APPLICANT: Gray, Jeff  
 APPLICANT: Lonberg, Nils  
 APPLICANT: Biosite Diagnostics, Inc.  
 APPLICANT: Genpharm International  
 TITLE OF INVENTION: Human Antibodies  
 FILE REFERENCE: 020015-00011005  
 CURRENT APPLICATION NUMBER: US/09/453,234

CURRENT FILING DATE: 1999-12-01  
 PRIOR APPLICATION NUMBER: US 60/157,415  
 PRIOR FILING DATE: 1999-10-02  
 NUMBER OF SEQ ID NOS: 112  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO: 36  
 LENGTH: 224  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-453-234-36

## Alignment Scores:

Pred. No.: 2.71e-71 Length: 224  
 Score: 1029.00 Matches: 201  
 Percent Similarity: 95.77% Conservative: 3  
 Best Local Similarity: 94.37% Mismatches: 9  
 Query Match: 59.00% Indels: 0  
 DB: 12 Gaps: 0

US-09-806-276a-3 (1-962) x US-09-453-234-36 (1-224)

```

QY 114 GAAATGTTGACACAGTCCTCCAGCCCTGCTTTGTCCTCCAGGGAAAGACCACC 173
   |||||
Db 1 GluileValLeuThrGlnSerProAlaThrLeuSerLeuSerProGlyLysArgAlaThr 20
QY 174 CTCCTCTGACAGGCGCAGTCAGAGTGTAGCAGCTTACCTGTGACCAAGAAACCT 233
   |||||
Db 21 LeuSerCysArgAlaSerGlnSerValSerSerTyrLeuAlaIlePtyrGlnGlnLysPro 40
QY 234 GGCAGGCTCCAGGCTCTCTCATATGATGCATCCAAAGGCGCAGTGGCATCCACC 293
   |||||
Db 41 GlyLysAlaProArgLeuLeuIleTyrAspAlaSerAsnArgAlaThrGlyIleProAla 60
QY 294 AGGTTACAGTGCAGTGGTCTGGAGACAGCTTACCTCACCACATCAGACAGACTGGAGCC 353
   |||||
Db 61 ArgPheSerGlySerGlySerGlyThrAspPheThrLeuThrIleSerSerLeuGlnPro 80
QY 354 GAAAGTGTGACACTTATTACTGTACAGAAATATTTTACTACTCGTACACTTTGGCCAG 413
   |||||
Db 81 GluAspPheAlaValIleTyrCysGlnGlnArgThrAsnTrpProArgThrPheGlyGln 100
QY 414 GGGACAGGCTGGAGATCAAGAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 473
   |||||
Db 101 GlyThrLysValGluIleLysArgThrValAlaAlaProSerValPheIlePheProPro 120
QY 474 TCTGATGACAGCTGAAATCTGGAATGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTAT 533
   |||||
Db 121 SerAspGlnGlnLeuLysSerGlyThrAlaSerValIleCysLeuLeuAsnAsnPheTyr 140
QY 534 CCCAGAGAGGCGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 593
   |||||
Db 141 ProArgLysAlaLysValGlnTrpLysValAlaAspAsnAlaLeuGlnSerGlyAsnSerGln 160
QY 594 GAGAGTGTACAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAG 653
   |||||
Db 161 GluSerValThrGlnGlnAspSerLysAspSerThrTyrSerLeuSerSerThrLeuThr 180
QY 654 CTGAGCAAGCAGACTACGAGAAACACAAAGTCTACCGCTGCGAAGTCAACCCATCAGGCG 713
   |||||
Db 181 LeuSerLysAlaAspTyrGlnLysLysIleValTyrAlaCysGlnValThrHisGlnGly 200
QY 714 CTGAGCTGCGCGCTCAAAAGAGCTTCAACAGGGGAGAG 752
   |||||
Db 201 LeuSerSerProValThrLysSerPheAsnArgGlyGln 213

```

## RESULT 12

US-10-216-484-52  
 Sequence 52, Application US/10216484  
 Publication No. US20030103976A1  
 GENERAL INFORMATION:  
 APPLICANT: Serizawa, No. US20030103976A1  
 APPLICANT: Serizawa, No. US20030103976A1  
 APPLICANT: Haryama, Hideyuki  
 APPLICANT: Nakahara, Kaori

Alignment Scores:	
Pred. No.:	2,73e-71
Score:	1029.00
Percent Similarity:	89.08%
Best Local Similarity:	85.29%
Query Match:	59.00%
DB:	15
Length:	238
Matches:	203
Conservative:	9
Mismatches:	22
Indels:	4
Gaps:	1

238  
203  
9  
22  
4  
1

QY 702 ACCCATCAGGGCCGTGAGCTCGCCCCGCACAAAGAGCTTCACACGGGGAGAGTGT 755  
|||||  
Db 221 TThHisGInGIyLeuSerSerProValThrLysSerPheAsnArgIylgIucYs 238

; Sequence 109, Application US/10216484  
; Publication No. US20030103976A1

GENERAL INFORMATION:  
APPLICANT: Serizawa, No. US20030103976A1ufusa

APPLICANT: Haruyama, Hidey  
APPLICANT: Nakahara, Kaori  
APPLICANT: Seizawa, No: 03

APPLICANT: Nakahara, Kaori  
APPLICANT: Tamaki, Ikuko  
APPLICANT: Tamaki, Ikuko

APPLICANT: Takahashi, Tohru  
TITLE OF INVENTION: Anti-Fas Antibodies

FILE REFERENCE: 980126CIP/HG  
CURRENT APPLICATION NUMBER: US/10/216,484

CURRENT AFFILIATION NUMBER: 00/10/2207  
 CURRENT FILING DATE: 2002-08-09  
 PRIOR APPLICATION NUMBER: US/09/499,662

;; PRIOR APPLICATION NUMBER: US/09/499,082  
;; PRIOR FILING DATE: 2000-02-09  
PRIOR APPLICATION NUMBER: US 09/053,583

; PRIOR APPLICATION NUMBER: US 09/053,583  
; PRIOR FILING DATE: 1998-04-01

```

;      NUMBER OF SEQ ID NOS: 165
;
; SEQ ID NO 109

```

```

; SEX ID NO 10
; LENGTH: 238
; TYPE: PRT

```

```

;
; TYPE: PRI
; ORGANISM: Artificial Sequence
FEATURE:

```

```

; FEATURE: Description of Artificial Sequence: Designed Light
; OTHER INFORMATION: chain of humanized anti-Fas antibody
; REMARK: INFORMATION:

```

OTHER INFORMATION;  
US-10-216-484-109

### Alignment Scores

alignement de  
Pred. No.:  
Score:

Score:  
Percent Similar  
Best Local Similar

Best Local Siml  
Query Match:

```
DB: 15 Gaps: 10 310 404 100 (1-328)
```

US-09-806-276A-3 (1-962) x US-10-216-484-109 (1-238)

54 ATGAAGCCCAAGCTCAGCTTCTCTTCTCTCTGCTACTCTGTGCTCCAGATVACCAACCGGA 113

Db 1 MetcIuThrAspThrIleLeuLeuTrpValLeuLeuLeuTrpValproGlySerThrgly 20

QY 114 GAATGTGTTGACACAGTCTCCAGCACCCTGTCTTTGTCTCCAGGGGAAGAGCCACC 173

Db 21 GlutIValLeuThrIInsPProGlyThrLeuSerLeuSerProGlyGluArgAlaThr 40

174 CTCCTCCGACAGGGCCACTCAGACTGT-----ACGAGCTACTTAGCCTGGTAC 221

41 leuercvysalaserGlnserValaspTyraspGlyaspserTyrmetasntptyr 60  
 42 .....  
 43 .....  
 44 .....  
 45 .....  
 46 .....  
 47 .....  
 48 .....  
 49 .....  
 50 .....  
 51 .....  
 52 .....  
 53 .....  
 54 .....  
 55 .....  
 56 .....  
 57 .....  
 58 .....  
 59 .....  
 60 .....  
 61 .....  
 62 .....  
 63 .....  
 64 .....  
 65 .....  
 66 .....  
 67 .....  
 68 .....  
 69 .....  
 70 .....  
 71 .....  
 72 .....  
 73 .....  
 74 .....  
 75 .....  
 76 .....  
 77 .....  
 78 .....  
 79 .....  
 80 .....  
 81 .....  
 82 .....  
 83 .....  
 84 .....  
 85 .....  
 86 .....  
 87 .....  
 88 .....  
 89 .....  
 90 .....  
 91 .....  
 92 .....  
 93 .....  
 94 .....  
 95 .....  
 96 .....  
 97 .....  
 98 .....  
 99 .....  
 100 .....

222 CAACAGAAACCTGGCCAGGCTCCAGGCTCCATCATATGATGACATCCACACAGGCCACT\* 281  
Db 41 leuercyslYsalasehnmsei valaspliyaspolyspoeelliymerchmnaap.y. 00

QY 222 CAACAGAAACTGGCCAGGCTCCATCATGATGATGCATCAACACCGGCCCTC  
|||||  
|||

Db 61 GINGInysProGIyGInAlaProArGLeuLeuIleTyAlaAlaIAsArAsnLeuGIIsEr 80

282 GGCATCCACCCAGGTTTCAGTGGCAGTGGTCTGGGACACAGCTTCACCTCACCATCAGC 341

Db 81 GlyIleProAspArgPheSerGlySerGlyLysThrAspPheThrLeuThrIleHis 1000

342 AGACTGGAGCCCGAAGATGTGGCACTTATATTACTGTCAGCAATATTTACTACTCCGTAC 401

Db 101 ProValGluGluGluAspAlaAtThrTyrTyrCysGlnGlnIserAsnGluAspProArg 120

402 ACTTTGGCCAGGGACGAGCTGGAGATCAACGAACTGTGGCTGCACCATCTGTCTTC 461

Db 121 Thphegi VG ng VTThcLysLeuG IjleLysArqrIrhValMaalaBProSeValPie 140

121 ThrPheGlyGlnGlyThrLysLeuGluIleLysAlaGlnIleValAlaAlaAlaIsoleuValThrLeu 521

462 ATCTTCCGCCCATCTGATGAGCAGTTGAATCTGGAACTGCCCTGTGTTGTGGCTGCG 521



```
QY 291 CCCAGGTTGAGTGGAGTGGCTGGGACAGACTTCTCAGCATCAGACAGACTGAG 350
    |||||||
Db 61 AsparGPheserGlyserGlyThrAspPheThrLeuThrIleSerArgLeuGlu 80
QY 351 CCCGAAGATGGCAGCTTTATTTACTGTACGCAATATTACTACT---CCGTACACTTT 407
    |||||||
Db 81 ProGlnuaspheIaValTyrTyrCysGlnGlnTyrGlySerSerProTyrThrPhe 100
QY 408 GGGCAGGGGACAGGCTGGAGATCAAGCAACTGTGGCTGACCATCTGCTTCATCTTC 467
    |||||||
Db 101 GlyGlnGlyThrLysLeuGlnIleLysArgThrValAlaIleProSerValPheIlePhe 120
QY 468 CCGCCATCTGATGAGCAGTGAATCTGAACTGCTCTGTGTGTGCTGCTGAATTAAC 527
    |||||||
Db 121 ProProSeraspGlnGlnLeuLysSerGlyThrAlaSerValValCysLeuLeuAsnAsn 140
QY 528 TTCTATCCAGAGAGGCCCAAGTAACAGTGAAGTGGATGAGCCCTCCCAATGGGTAAC 587
    |||||||
Db 141 PheTyrProArgGlnAlaLysValGlnTyrPargValAlaSpasnAlaLeuGlnSerGlyAsn 160
QY 588 TCCAGAGAGAGTGTACAGAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 647
    |||||||
Db 161 SerGlnGlnSerValThrGlnGlnAspSerLysaspSerThrTyrSerLeuSerSerThr 180
QY 648 CTGACGCTGAGCAAGCAGACTACGAGAAACACAAGCTTACGCTGCGAAGTCACCCAT 707
    |||||||
Db 181 LeuThrLeuSerLysAlaAspTyrGlnLysHisLysValTyrAlaCysGlnValThrHis 200
QY 708 GAGGGCTGAGCTGCGCCGTCACAAGAGCTTCAACAGGGGAGAG 752
    |||||||
Db 201 GlnGlyLeuSerSerProValThrLysSerPheAsnArgGlyGlu 215
```

Search completed: July 18, 2003, 17:44:17  
Job time : 41.9242 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: July 18, 2003, 17:15:57 ; Search time 29.0684 Seconds

(without alignments)  
6363.010 Million cell updates/sec

Title: US-09-806-276a-3

Perfect score: 1744

Sequence: 1 tcgagcgcgattgcgctcgag.....gaatttcgcaaaaaaaaaa 962

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Dgapop 6.0	Dgapext 7.0
Delop 6.0	Delext 7.0

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=xlp  
-O=/cgn2.1/USPRO.spool/US09806276/runat\_18072003.160925.12767/app.query.fasta.1.3854  
-DB=PIR\_73 -QFMT=fastan -SUFFIX=ipr -MINMATCH=0.1 -LOOPEL=0 -LIST=45  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45  
-DOCCALIGN=200 -THR.SCORE=Pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09806276.ecgn.1.1.185.ecgnat.18072003.160925.12767 -NCPu=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1018.5	58.4	215	JE0242	Ig kappa chain NIG
2	993.5	57.0	215	JE0244	Ig kappa chain NIG
3	974.5	55.9	215	A23746	Ig kappa chain V-I
4	956.5	54.8	215	JE0243	Ig kappa chain NIG
5	899	51.5	216	JE0241	Ig kappa chain Am3
6	806	46.2	240	SG0684	Ig kappa chain pre
7	741	42.5	234	SG1320	Ig kappa chain pre
8	740	42.4	234	SG14237	Ig kappa chain pre
9	736.5	42.2	230	SG3161	Ig kappa chain - s
10	726.5	41.7	225	SG7484	Ig kappa chain - m
11	717	41.1	218	SG8241	Ig kappa chain - m
12	712.5	40.9	219	S38865	Ig kappa chain - m
13	712	40.8	220	A31790	Ig kappa chain V r
14	711	40.8	218	UC5810	Ig kappa chain V r

15	706	40.5	214	SG8212	Ig kappa chain (Ma
16	703.5	40.3	219	SG2028	Ig kappa chain - m
17	703.5	40.3	235	SG2058	Ig kappa chain - m
18	700.5	40.2	219	PC4203	Ig kappa chain (mo
19	700.5	40.2	219	SG1112	Ig kappa chain V r
20	698.5	40.1	217	SG2772	Ig kappa chain - m
21	693	39.7	210	A56169	Ig kappa chain V r
22	680.5	39.0	225	JI0029	Ig kappa chain pre
23	671	38.5	144	PI0106	Ig kappa chain pre
24	613.5	35.2	178	SG20631	Ig kappa chain - h
25	613	35.1	145	PT0219	Ig kappa chain V-C
26	610.5	35.0	134	SG8643	Ig kappa chain V r
27	603.5	34.6	197	SG2953	Ig kappa chain V r
28	602	34.5	135	SG2059	Ig kappa chain (WM
29	589.5	33.8	129	K3H0HA	UC-kappa protein -
30	589.5	33.8	229	A20969	Ig kappa chain pre
31	584.5	33.5	129	SG4932	Ig kappa chain pre
32	583.5	33.5	129	K3H0HI	anti-Sm antibody V
33	581.5	33.3	129	SG4369	Ig kappa chain pre
34	580	33.3	128	SG40379	Ig kappa chain var
35	577.5	33.1	128	SG20636	Ig kappa chain V-J
36	576	33.0	128	A56701	Ig kappa chain V r
37	573	32.9	128	K3H0U1	Ig kappa chain pre
38	572.5	32.8	129	SG40363	Ig kappa chain - h
39	572.5	32.8	128	A49633	Ig kappa chain - h
40	571.5	32.8	128	A32274	Ig kappa chain pre
41	567	32.5	128	SG40343	Ig kappa chain V-J
42	559	32.1	125	SG40344	Ig kappa chain V-J
43	558	32.0	129	SG2627	Ig kappa chain V r
44	557.5	32.0	130	SG2637	Ig kappa chain V r
45	555.5	31.9	121	SG40327	Ig kappa chain - h

#### ALIGNMENTS

RESULT 1  
JE0242  
Ig kappa chain NIG26 precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 05-Dec-1998 #sequence.revision 05-Dec-1998 #text.change 21-Jan-2000  
C:Accession: JE0242  
R:Alim, M.A.; Yamaki, S.; Hossain, M.S.; Takeda, K.; Kojima, M.; Takashi, I.; Shinoda  
submitted to JIPID, November 1998  
A:Description: Structure relationship of kappa type light chains with AL amyloidosis:  
A:Reference number: JE0241  
A:Accession: JE0242  
A:Molecule type: protein  
A:Residues: 1-215 <ALI>  
C:Superfamily: immunoglobulin V region; immunoglobulin homology  
F:16-91/Domain: immunoglobulin homology <IMM>

#### Alignment Scores:

Pred. No.:	2.98e-59	Length:	215
Score:	1018.50	Matches:	199
Percent Similarity:	95.35%	Conservative:	6
Best Local Similarity:	92.56%	Mismatches:	9
Query Match:	58.40%	Indels:	1
DB:	2	Gaps:	1

US-09-806-276a-3 (1-962) x JE0242 (1-215)  
QY 114 GAATGTGTGACACAGCTCCAGCCACCCGCTGCTTCCTCCAGGGGAAGAGCCACC 173  
Db 1 GlttlevallleuthrglnserProglyThrleuserleuserProglygluThrgalThr 20  
QY 174 CTCCTCCGACGGCCAGCTCAGAGTGTAGC---AGCTACTTACCTCGTACCAACAGAA 230  
Db 21 leusercysArgrlaSergrlnserValserAsnAsnTyrleuAlarpryrglnGlnlys 40  
QY 231 CCTGGCCAGGCTCCAGAGCTCCTCATCTATGATGATTCACACAGGGCCAGGCATCCCA 290  
Db 41 ProglyGlnAlarProserleuserleuserleuserYAspAlaserSerArgrlaThrGlylePro 60

```

QY 291 CCCAGTTCAGTGGCAGATGGGCTGTGGACAGACTTCACTCTCCATCCAGACAGCTGGAG 350
    |||||||
Db 61 Aspargheserilyserglyserglythraspheiileuthrilleserglyleuglu 80
QY 351 CCCGAGATGTGGCAGCTTTTACTGTCTGACGACATATTTTACTCTCCGACATCTTTGGC 410
    |||||||
Db 81 Proglunaphenalaivaltyrtyrcysglnghlntyrasparproprothrphlegly 100
QY 411 CAGGGGACAGGCTGGAGATCAACAGACTGTGGCTGCACCATCTGTCTTCATCTTCCG 470
    |||||||
Db 101 Glnghlythrlysvaigluillelysarqthvalalalaproservalpheiilepro 120
QY 471 CCATGTATGAGAGCTGAATCTGGAATCTGGAATCTGTTGTGTGCTGCGAATTAATCTTC 530
    |||||||
Db 121 ProseraspeluglnleuylsserglythraliserValcysleuenaasnph 140
QY 531 TATCCAGAGAGGCCAAGATACAGTGGAGGTGATACGCCCTCCATCGGGTAACTCC 590
    |||||||
Db 141 Tyrproargluinalalysvalglntrplyvalaspsnalaleuinserglyanser 160
QY 591 CAGGAGAGTGTCCACAGCAGACAGACAGACAGACAGACAGACAGACAGACAGACAG 650
    |||||||
Db 161 GlnghlyserValthrlyglnaspserylsserthryrserleuserSerThrleu 180
QY 651 ACCGTGACCAAGCAGACTACGAGAAACACAAAGTCTACGCTGCGAAGTCAACCCATCAG 710
    |||||||
Db 181 Thrlleuserlysalasprtyglulyshtlysvallyralacysgluvalthrhisgln 200
QY 711 GGCTGAGCTGCCCGCTCACAAGAGCTTCAACAGGGGAGAGTGT 755
    |||||||
Db 201 GlyleuserSerProvalThrlyserPheasnarglylucys 215

```

## RESULT 2

```

JEO244
Ig kappa chain NIC2 precursor - human
C:Species: Homo sapiens (man)
C:Date: 05-Dec-1998 #sequence.revision 05-Dec-1998 #text_change 21-Jan-2000
C:Accession: JEO244
R:Allm, M.A.; Harz, Y.; Hosseini, M.S.; Takeda, K.; Yamagata, F.; Yamaki, S.; Kazi, H.; T
submitted to JRPID, November 1998
A:Description: A new subgroup of k type light chains (VK) identified in cases of AL amy
A:Reference number: JEO243
A:Accession: JEO244
A:Molecule type: protein
A:Residues: 1-215 <ALI>
C:Superfamily: immunoglobulin V region; immunoglobulin homology
F:16-90/Domain: immunoglobulin homology <IMM>

```

## Alignment Scores:

```

Pred. No.: 1,29e-57 Length: 215
Score: 993.50 Matches: 196
Percent Similarity: 93.49% Conservative: 5
Best Local Similarity: 91.16% Mismatches: 13
Query Match: 56.97% Indels: 1
DB: 2 Gaps: 1

```

US-09-806-276a-3 (1-962) x JEO244 (1-215)

```

QY 114 GAAATGTGTGACACAGTCTCCAGCACCCCTGTCTTGTCTCCAGGGGAAAGGCCACC 173
    |||||||
Db 1 GluValValleuthrlnghlnserProialthrleuserValserProglygluargalathr 20
QY 174 CTCCTCTGAGGGCAGTACAGAGTGTAGCAGTACTTACCTGTGTACCAAGAAAGTCT 233
    |||||||
Db 21 LeuserlysalgalaserGlnserValHisserasnleuathrprtyrglnghlnlypro 40
QY 234 GCGCAGGCTCCAGGCTCTCATCTATGATGATCATCAACAGGGCCACTGCATCCACCC 293
    |||||||
Db 41 GlyGlnalaproargleuileuthrlyrargalaserThrargalathrghlylleproala 60
QY 294 AGGTTCAAGTGGATGGTGTGGGACAGACTTCACTCTCCACATCAGACAGCTGAGACC 353
    |||||||
Db 61 ArghpserglyserglyserglythraspheiileuthrilleSerSerleuglnser 80

```

```

QY 354 GAAGATGTGGCAGCTTTTACTGTGACGACATATTTTACT---ACTCGTACACTTTTGGC 410
    |||||||
Db 81 Gluasphealaleuthrtyrtyrcysglnghlntyrasnthrtrpproproleuthrphlegly 100
QY 411 CAGGGGACAGCTGGAGATCAACAGACTGTGGCTGCACCATCTGTCTTCATCTTCCG 470
    |||||||
Db 101 GlyglythrlysvaigluillelysarqthvalalalaproserValpheiilepro 120
QY 471 CCATGTATGAGAGCTGAATCTGGAATCTGGAATCTGTTGTGTGCTGCGAATTAATCTTC 530
    |||||||
Db 121 ProseraspeluglnleuylsserglythraliserValcysleuenaasnph 140
QY 531 TATCCAGAGAGGCCAAGATACAGTGGAGGTGATACGCCCTCCATCGGGTAACTCC 590
    |||||||
Db 141 Tyrproargluinalalysvalglntrplyvalaspsnalaleuinserglyanser 160
QY 591 CAGGAGAGTGTCCACAGCAGACAGACAGACAGACAGACAGACAGACAGACAGACAG 650
    |||||||
Db 161 GlnghlyserValthrlyglnaspserylsserthryrserleuserSerThrleu 180
QY 651 ACCGTGACCAAGCAGACTACGAGAAACACAAAGTCTACGCTGCGAAGTCAACCCATCAG 710
    |||||||
Db 181 Thrlleuserlysalasprtyglulyshtlysvallyralacysgluvalthrhisgln 200
QY 711 GGCTGAGCTGCCCGCTCACAAGAGCTTCAACAGGGGAGAGTGT 755
    |||||||
Db 201 GlyleuserSerProvalThrlyserPheasnarglylucys 215

```

## RESULT 3

```

A23746
Ig kappa chain V-III (KAU cold agglutinin) - human
C:Species: Homo sapiens (man)
C:Date: 30-Dec-1991 #sequence.revision 30-Dec-1991 #text_change 21-Jan-2000
C:Accession: A23746
R:Leonl, U.; Ghiso, J.; Goni, F.; Frangione, B.
J. Biol. Chem. 266, 2856-2842, 1991
A:Title: The primary structure of the Fab fragment of protein KAU, a monoclonal immun
A:Reference number: A23746; MUID:91131575; PMID:1993660
A:Accession: A23746
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-215 <LDE>
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heteroleptamer; immunoglobulin
F:16-91/Domain: immunoglobulin homology <IMM>

```

## Alignment Scores:

```

Pred. No.: 2,26e-56 Length: 215
Score: 974.50 Matches: 197
Percent Similarity: 94.86% Conservative: 6
Best Local Similarity: 92.06% Mismatches: 10
Query Match: 55.88% Indels: 1
DB: 2 Gaps: 1

```

US-09-806-276a-3 (1-962) x A23746 (1-215)

```

QY 114 GAAATGTGTGACACAGTCTCCAGCACCCCTGTCTTGTCTCCAGGGGAAAGGCCACC 173
    |||||||
Db 1 GluileValleuthrlnghlnserProialthrleuserSerleuserProglygluargalathr 20
QY 174 CTCCTCTGAGGGCAGTACAGAGTGTAGCAGC---TACTTACCTGTGTACCAACAGAA 230
    |||||||
Db 21 LeuserglyllyalaserGlnserValSerSerasnlyrleuvalathrprtyrglnghlnlys 40
QY 231 CCGGCGCAGGCTCCAGGCTCTCATCTATGATGATCATCAACAGGGCCAGTGGATCCCA 290
    |||||||
Db 41 ProglyGlnalaproargleuileuthrlyrargalaserThrargalathrghlyllepro 60
QY 291 CCCAGTTCAGTGGCAGTGGTGTGGGACAGACTTCACTCTCCACATCAGACAGACTGAGAG 350
    |||||||
Db 61 AspargheserilyserglyserglythraspheiileuthrilleSerargleu 80

```

```

QY 351 CCCGAGATGTCGACCTTATTACTGTCAGCAATATTTACTCCGTACACTTTGGC 410
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 81 ProGluAspPheAlaValTyrTrpGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
QY 411 CAGGGAGCAGGCTGGAATCAACGAGACTGGCTGCACCATCTGCTTCTTCCCG 470
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 101 GlyThrLysValGlnIleLysArgThrValAlaAlaProSerValPheIlePhePro 120
QY 471 CCATCTGATGACAGTGAATCTGGAACCTGCTCTGTGTGGCTGGCTCAATATCTC 530
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 121 ProSerAspGlnGlnIleLysSerGlyThrAlaSerValValGlyLeuLeuAsnAspHe 140
QY 531 TATCCAGAGAGGCCCAAGTACAGTGAAGGTGATAAGCCCTCAATCCGCTAATCC 590
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 141 TyrProArgGlnAlaLysValGlnIleTrpLysValAspAsnAlaLeuGlnIleSer 160
QY 591 CAGGAGAGTGCACAGACGACGACAGACGACGACCTACAGCTCAGACGACCTG 650
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 161 GlnGlnSerValThrGlnIleAspSerLysAspSerThrThrLysSerLysSerThrLeu 180
QY 651 ACCGTGACCAAGACGACTACAGAAACACAAAGTACAGCTCAGAGTACACCATCAG 710
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 181 ThrLeuSerLysAlaAspTyrGlnLysHisLysValTyrAlaGlyLeuAlaThrHisGln 200
QY 711 GGCTGAGCTGCGCCGCTCACAAGAGCTTCAACAGGGAGAG 752
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 201 GlyLeuSerSerProValThrLysSerPheAsnArgGlyGlu 214

```

## RESULT 4

```

JE0243
Ig kappa chain N193 precursor - human
C:Species: Homo sapiens (man)
C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jan-2000
C/Accession: JE0243
R:Alim, M.A.; Hara, Y.; Hossain, M.S.; Takeda, K.; Yamagata, F.; Yamaki, S.; Kazi, H.; T
submitted to JIPID, November 1998
A:Description: A new subgroup of k type light chains (VK) identified in cases of AL amy
A:Reference number: JE0243
A:Accession: JE0243
A:Molecule type: protein
A:Residues: 1-215 <ALI>
C:Superfamily: immunoglobulin V region; immunoglobulin homology
F:16-90/Domain: immunoglobulin homology <IMM>

```

```

Alignment Scores:
Pred. No.: 3.41e-55 Length: 215
Score: 956.50 Matches: 187
Percent Similarity: 92.09% Conservative: 11
Best Local Similarity: 86.98% Mismatches: 16
Query Match: 54.85% Indels: 1
DB: 2 Gaps: 1

```

```

US-09-806-276a-3 (1-962) x JE0243 (1-215)
QY 114 GAAATTGTGTGACACAGTCTCAGCCACCCCTGTTGTCCTCCAGGGGAAAGCCACC 173
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1 GlnIleValMetThrGlnSerProAlaThrLeuSerValSerProGlyGluArgAlaThr 20
QY 174 CTCTCTGCAAGGCCAGTCAAGTGTGAGCTTACCTGCTGATCAACAGAAACCT 233
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 21 LeuSerCysArgAlaSerGlnSerValAlaThrAsnValValTrpLysMetGlnLysLeu 40
QY 234 GGCAGGCTCCCAAGGCTCCATCATATGATGATCCAGAGGCCAGCTGCGATCCACCC 293
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 41 GlnGlnIleProArgLeuLeuIleTyrAspAlaSerThrArgAlaThrGlyValProAla 60
QY 294 AGGTCAGTGGCAGTGGTCTGGAGACAGTTCACCTCAGCTCAGAGCTGGAGGCC 353
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 61 ArgPheSerGlySerGlySerGlyThrGlnPheThrLeuThrLysSerLeuGlnSer 80
QY 354 GAAGATGTGGCACTTATTACTGTGACGAAATATTTTACTACTCCGTACACTTTGGCCAG 413
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 81 GluAspPheAlaIleTyrTrpCysGlnHisAsnAlaIleTrpProThrPheGlyGln 100

```

```

QY 414 GGGACCGAGCTGGAGATCAAA---CGAGCTGGCTGCACCAATCTCTTCACTTCCCG 470
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 101 GlyThrLysValGlnIleThrLysArgThrValAlaAlaProSerValPheIlePhePro 120
QY 471 CCATCTGATGACAGTGAATCTGGAACCTGCTCTGTGTGGCTGGCTCAATATCTC 530
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 121 ProSerAspGlnGlnIleLysSerGlyThrAlaSerValValCysLeuLeuAsnAspHe 140
QY 531 TATCCAGAGAGGCCCAAGTACAGTGAAGGTGATAAGCCCTCAATCCGCTAATCC 590
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 141 TyrProArgGlnAlaLysValGlnIleTrpLysValAspAsnAlaLeuGlnIleSer 160
QY 591 CAGGAGAGTGCACAGACGACGACACAGACGACGACCTACAGCTCAGACGACCTG 650
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 161 GlnGlnSerValThrGlnIleAspSerLysAspSerThrThrLysSerLysSerThrLeu 180
QY 651 ACCGTGACCAAGACGACTACAGAAACACAAAGTACAGCTCAGAGTACACCATCAG 710
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 181 ThrLeuSerLysAlaAspTyrGlnLysHisLysValTyrAlaCysGlnValThrHisGln 200
QY 711 GGCTGAGCTGCGCCGCTCACAAGAGCTTCAACAGGGAGAGTGT 755
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 201 GlyLeuSerSerProValThrLysSerPheAsnArgGlyGluCys 215

```

## RESULT 5

```

JE0241
Ig kappa chain Am37 precursor - human
C:Species: Homo sapiens (man)
C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jan-2000
C/Accession: JE0241
R:Alim, M.A.; Yamaki, S.; Hossain, M.S.; Takeda, K.; Kojima, M.; Takashi, I.; Shinoda
submitted to JIPID, November 1998
A:Description: Structure relationship of kappa type light chains with AL amyloidosis:
A:Reference number: JE0241
A:Accession: JE0241
A:Molecule type: protein
A:Residues: 1-216 <ALI>
C:Superfamily: immunoglobulin V region; immunoglobulin homology
F:16-92/Domain: immunoglobulin homology <IMM>

```

```

Alignment Scores:
Pred. No.: 1.98e-51 Length: 216
Score: 899.00 Matches: 177
Percent Similarity: 89.91% Conservative: 19
Best Local Similarity: 81.19% Mismatches: 16
Query Match: 51.35% Indels: 6
DB: 2 Gaps: 3

```

```

US-09-806-276a-3 (1-962) x JE0241 (1-216)
QY 114 GAAATTGTGTGACACAGTCTCAGCCACCCCTGTTGTCCTCCAGGGGAAAGCCACC 173
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1 AspIleValLeuThrGlnSerProAspPheLeuAlaValSerLeuGlyGluArgAlaThr 20
QY 174 CTCTCTGCAAGGCCAGTCAAGTGTGAGCTTACCTGCTGATCAACAGAAACCT 233
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 21 IleAsnCysLysSerSerGlnSerValLeuTyrAsnSerLysAsnPheLeuAlaTrpLys 40
QY 222 CAACGAAACCTGGCCAGGCTCCAGGCTCTCATATGATGATCCAGAGGCCAGCT 281
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 41 GlnGlnLysProGlyGln---ProLysLeuLeuIleTrp---AlaAsnValArgLysSer 58
QY 282 GGCATCCCAACAGGCTGAGTGGCAGGCTGGGTCGGGACAGCTTCACTCAGCATCAGC 341
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 59 GlyValProAspArgPheArgGlySerGlyValAlaGlyThrAspPheThrLeuThrLys 78
QY 342 AGACTGAGGCCGAAGATGTGGCACTTATTACTGTGACGAAATATTTTACTACTCCGTAC 401
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 79 AsnLeuGlnAlaGlnLeuValAlaValTyrTrpCysGlnGlnIleTyrTrpThrProTyr 98
QY 402 ACTTTGGCCAGGGAGCAGGCTGGAAGATCAACAGAACTGTGCTGACCATCTGTCTTC 461
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Db 99 SerPheGlyInGlyIleValGluIleValArgThrValAlaIleProSerValPhe 118  
 Oy 462 ATCTTCCCGCCATCTGATGAGCAGTTGAAATGCAATGCTCTGTGTGCTGCTG 521  
 Db 119 IleHerProSerAspGluInGlyLeuLeuSerGlyThrAlaSerValValCysLeu 138  
 Oy 522 AATACTTCTATCCAGAGAGGCCAAAGTACAGTGAAGGTGATACGCCCTCCATG 581  
 Db 139 AsnAsnHeterProAlaGluAlaValGlnTrpLysValAlaPaspAlaLeuGlnSer 158  
 Oy 582 GGTAACGCCAGGAGTGTACAGAGCAGACAGCAGACAGCAGCAGCAGCAGCAGC 641  
 Db 159 GlySerSerGlnGlnSerValThrGlnGlnAspSerLysAspSerThrLysLeuSer 178  
 Oy 642 AGCAGCTGACGCTGACCAAGCAGACTACAGACAAACAAAGTCTACGCTCGAAGTC 701  
 Db 179 SerThrLeuThrSerLysAlaAspTrpGlnLysHisLysValTrpAlaCysGluVal 198  
 Oy 702 ACCATCAGGCGCTGAGCTGCGCCGTCACAAAGAGCTTCAAGGAGGAGTGT 755  
 Db 199 ThrHisGlnGlyLeuSerSerProValThrLysSerPheAsnArgGlyLucys 216

## RESULT 6

S06084  
 Ig kappa chain precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 21-Jan-2000  
 C:Accession: S06084  
 R:Crowe, J.S.; Smith, M.A.; Cooper, H.J.  
 Nucleic Acids Res. 17, 7992, 1989  
 A:Title: Nucleotide sequence of Y3-Ag 1.2.3. rat myeloma immunoglobulin kappa chain cDNA  
 A:Reference number: S06084; MUID:90016888; PMID:2508067  
 A:Accession: S06084  
 A:Molecule type: mRNA  
 A:Residues: 1-240 <CRO>  
 A:Cross-references: EMBL:X16129; NID:956457; PIDN:CAA34256.1; PID:956458  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-240/Product: Ig kappa chain #status predicted <MAT>  
 F:153-222/Domain: immunoglobulin homology <IMM>

## Alignment Scores:

Pred. No.:	2,4e-45	Length:	240
Score:	806.00	Matches:	149
Percent Similarity:	78.33%	Conservative:	39
Best Local Similarity:	62.08%	Mismatches:	46
Query Match:	46.22%	Indels:	6
DB:	2	Gaps:	1

US-09-806-276a-3 (1-962) x S06084 (1-240)

Oy 54 ATGAGAGCCCGCCAGCTCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113  
 Db 1 MetGlnSerGlnThrGlnValLeuMetSerLeuLeuLeuTrpIleSerGlyThrCysGly 20  
 Oy 114 GAAATGTGTGACAGCTTCACAGCACCCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 173  
 Db 21 AspPheValMetThrGlnSerProSerSerLeuAlaValSerAlaGlyGlnThrValThr 40  
 Oy 174 CTCTCTGAGGCGCCAGTTCAGAGTGT-----AGCAGCTACTTACCC 215  
 Db 41 IleAsnGlySerSerSerGlnSerLeuPheTrpSerGlyAsnGlnLysAsnTrpLeuAla 60  
 Oy 216 TGGTACCAACAAACCTGGCCAGGCTCCAGGCTCCACCTCATCTATGATGACCAACAGG 275  
 Db 61 TrpTyrGlnGlnLysProGlnSerProLysLeuLeuLysTrpPalaSerThrArg 80  
 Oy 276 GGCAGGAGTCCAGCCAGCTTCAGTGCAGTGGTGGTGGAGACAGCTTCACTTCACTCACC 335  
 Db 81 GlnSerGlyValProAspArgPheIleGlySerGlySerGlyThrAspPheThrLeuThr 100  
 Oy 336 ATCAGCAGAGCTGAGCCCGCAAGATGTGCACTTATTACTGTACAGCAATATTACTACT 395

Db 101 IleSerSerValGlnAlaGlnAspLeuAlaLeuTrpTrpCysLeuGlnTrpGluThr 120  
 Oy 396 CCGTACACTTTTGGCCAGGAGGAGCAGGCTGAGATCAACAGACAGTGGCTGCACATCT 455  
 Db 121 ProTyrThrPheGlyAlaGlyThrLysLeuGlnLeuLysArgAlaAspAlaIleProThr 140  
 Oy 456 GTCTTCATCTTCCCGCCATCTGATGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 515  
 Db 141 ValSerIlePheProProSerThrGlnGlnLeuAlaTrpGlyAlaSerValValCys 160  
 Oy 516 CTGCTGAATTAATCTTATCCAGAGAGCCCAAGATACAGTGAAGTGAAGTGAAGTGAAGT 575  
 Db 161 LeuMetAsnAsnPheTrpProArgAspLysSerValLysTrpLysIleAspGlyThrGln 180  
 Oy 576 CAATGGGTAATCTCCAGAGAGTGTTCACAGAGCAGCAGCAGCAGCAGCAGCAGCAGC 635  
 Db 181 ArgArgAspGlyValLeuAspSerValThrAspGlnAspSerLysAspSerThrLysSer 200  
 Oy 636 CTCAGCAGACCCCTGACGCTGACCAAGCAGACTACAGAGAAACAAAGTCTACGCCCTGC 695  
 Db 201 MetSerSerThrLeuSerLeuSerLysAlaAspTrpGlnSerHisAsnLeuThrCys 220  
 Oy 696 GAAGTCACCCATCAGGCGCTGAGCTGCGCCGTCACAAAGAGCTTCAACAGGAGAGTGT 755  
 Db 221 GluValValHisLysThrSerSerSerProValValLysSerPheAsnArgGlyLucys 240

## RESULT 7

S01320  
 Ig kappa chain precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 21-Jan-2000  
 C:Accession: S01320  
 R:de Waele, P.; Feys, V.; van de Voorde, A.; Molemans, F.; Fiers, W.  
 Eur. J. Biochem. 176, 287-295, 1988  
 A:Title: Expression in non-lymphoid cells of mouse recombinant immunoglobulin directe  
 A:Reference number: S01320; MUID:88329081; PMID:3138116  
 A:Accession: S01320  
 A:Molecule type: mRNA  
 A:Residues: 1-234 <DE1>  
 A:Cross-references: EMBL:X13187; NID:951784; PIDN:CAA31579.1; PID:951785  
 A:Note: This sequence was determined from the differentiated gene  
 C:Superfamily: Immunoglobulin V region; immunoglobulin homology  
 C:Keywords: heterotetramer; immunoglobulin  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-234/Product: Ig kappa chain #status predicted <MAT>  
 F:36-110/Domain: immunoglobulin homology <IMM>

## Alignment Scores:

Pred. No.:	4.33e-41	Length:	234
Score:	741.00	Matches:	136
Percent Similarity:	75.21%	Conservative:	40
Best Local Similarity:	58.12%	Mismatches:	58
Query Match:	42.49%	Indels:	0
DB:	2	Gaps:	0

US-09-806-276a-3 (1-962) x S01320 (1-234)

Oy 54 ATGAGAGCCCGCCAGCTCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113  
 Db 1 MetSerValProThrGlnValLeuGlyLeuLeuLeuLeuTrpLeuThrAspAlaArgCys 20  
 Oy 114 GAAATGTGTGACAGCTTCACAGCACCCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 173  
 Db 21 AspIleGlnMetThrGlnSerProAlaSerLeuSerValSerValGlnLysSerValThr 40  
 Oy 174 CTCTCTGAGGCGCCAGTTCAGAGTGTATGAGCTACTTATGCGCTGACCAAGAACT 233  
 Db 41 IleThrCysArgAlaSerGlnAsnLysTrpSerAsnLeuAlaTrpTrpGlnGlnLysGln 60  
 Oy 234 GGCAGGCTCCAGAGCTCTCATATATGATGATCCACAGGAGGAGCAGCTGACATCCACCC 293  
 Db 61 GlyLysSerProGlnLeuLeuValLysValAlaThrLysLeuValAspGlyValProSer 80





```

QY 186 GCCATCGAGTGTAGACCTACTTACCTGTGATCCACGAAACCTGGCCAGCTCC 245
    |||||
DB 41 ThrsGlnSerValSerAsnTyrLeuAsnTyrGlnGlnLysProGlyGlnAlaPro 60
QY 246 AGGCTCCATCATTAATGATCCACGAGGGCCCTGGATCCACCCAGGCTTACAGGCG 305
    |||||
DB 61 LysLeuLeuLeuTyrTyrAlaThrArgLeuHisThrAspValProSerArgPheSerGly 80
QY 306 AGTGGCTGTGGACGACTTCACTCTACATCAGACAGACAGAGCCGGAAGATGTGGCA 365
    |||||
DB 81 SerGlySerGlyThrAspTyrThrLeuThrIleSerAsnLeuGlnAlaAsnAspThrAla 100
QY 366 CTTTACTGTGACCATATTTTACTTACCTCCGATACACTTTGGCCAGGGGAGCAGGCTG 425
    |||||
DB 101 ThrTyrTyrCysLeuGlnTyrGlnSerThrProLeuAlaPheGlyGlyGlyThrAsnAl 120
QY 426 GAGATCAACAGCAAGCTGGCTGCACACCTATCTTCATCTTCCGCCATCTGATGAGAG 485
    |||||
DB 121 GlnIleLysArgSerAspAlaGlnProSerValPheLeuPheLysProSerGlnGln 140
QY 486 TTGAATCTGCACTGCTCTGTGTGTGCTGCTGTAATTAATCTTATCCAGAGAGGCC 545
    |||||
DB 141 LeuAlaGlnThrGlyThrValSerValValCysLeuValAsnAspPheTyrProLysAspIle 160
QY 546 AAGATACAGGTGAGAGGTGATTAAGCCCTCCATCGGGTAACTCCAGAGAGAGTGTCA 605
    |||||
DB 161 AsnValLysValLysValValAspGlyValThrGlnAsnSerAsnPheGlnAsnSerPheThr 180
QY 606 GAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 665
    |||||
DB 181 AspGlnAspSerLysLysSerThrTyrSerLeuSerSerThrLeuThrLeuSerSerSer 200
QY 666 GACTACGAGAACACAAAGTCTACGCTCCGCAAGTCAACCCATCAAGGCGCTAGCTCGGCC 725
    |||||
DB 201 GlnTyrGlnSerHisAsnAlaTyrValAcysGlnValSerHisSerLeuProThrAla 220
QY 726 GTCAACAAAGAGCTTCAACAGGAGGAGAGTGT 755
    |||||
DB 221 LeuValLysSerPheAsnLysAsnGlnLys 230

```

## RESULT 10

```

S37484
I: kappa chain - mouse (fragment)
C: Species: Mus musculus (house mouse)
C: Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 11-Jan-2000
C: Accession: S37484
R: Duncanson, F.F.D.
submitted to the EMBL Data Library, February 1993
A: Reference number: S37483
A: Accession: S37484
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-225 <CDUC>
A: Cross-references: EMBL:X70424; NID:q406254; PIDN:CAAA9869.1; PID:q406255
C: Superfamily: Immunoglobulin V region; Immunoglobulin homology
C: Keywords: heterotetramer; Immunoglobulin

```

## Alignment Scores:

```

Pred. No.: 3,87e-40 Length: 225
Score: 726.50 Matches: 137
Percent Similarity: 78.12% Conservative: 38
Best Local Similarity: 61.16% Mismatches: 48
Query Match: 41.66% Indels: 1
Gaps: 2

```

US-09-806-276a-3 (1-962) x S37484 (1-225)

```

QY 84 CTGCTACTGCTGCTCCAGATACCAACGAGAAATGTTGTGACAGCTTCCAGCCACC 143
    |||||
DB 3 LeuLeuLeuCysValSerLysLysLysLysLysLysLysLysLysLysLysLysLys 22
QY 144 CTGCTTTTGTCTCCAGGGAAGAGCCCTCTCTCGAGGGCCAGTCAAGATGTTAGC 203
    |||||

```

```

DB 23 LeuLeuLeuSerAlaLysAspArgValThrIleThrCysLysAlaSerGlnSerValSer 42
QY 204 AGCTACTTACCTGTGATACCAACAGAAACCTGGCCAGGCTCCAGGCTCTCTATGAT 263
    |||||
DB 43 AsnAspValAlaLysPyrGlnGlnLysProGlyGlnSerProLysLeuLeuLeuTyr 62
QY 264 GCATCCACAGGGCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCC 323
    |||||
DB 63 AlaSerSerArgTyrThrValProAspArgPheThrGlySerGlyTyrGlyThrAsp 82
QY 324 TTCACCTCCACATCAGACAGACTGAGAGCCGGAAGATGTGACATTATTTACTGTGACAA 383
    |||||
DB 83 PheThrPheThrIleSerThrValGlnAlaGlnAspLeuAlaValTyrPheCysGln 102
QY 384 TATTTTACTTACTCCGATACCTTTGGCCAGGGAGCAGGCTGAGATCAACAGACTGCG 443
    |||||
DB 103 AspTyrSerSer---TyrThrPheGlyGlyGlyThrLysLeuGlnIleLysArgAlaAsp 121
QY 444 GCTGCACCATCTGCTTCATCTTCCGCCATCTGATGAGCAGTGAATCTGAACTGCC 503
    |||||
DB 122 AlaAlaProThrValSerIlePheProProSerSerGlnGlnLeuThrSerGlyAla 141
QY 504 TCTGTGTGTGCTGCTGATTAATCTTATCCAGAGAGCCAAATCAGTGAAGTGT 563
    |||||
DB 142 SerValValCysPheLeuAsnAsnPheTyrProLysAspIleAsnValLysTrpLysIle 161
QY 564 GATACAGCCCTCCATTCGGTAACTCCAGAGAGTGTCAACAGCAGCAGCAGCAGCAGC 623
    |||||
DB 162 AspLysSerGlnLysGlnAsnGlyValLeuAsnSerThrPheAsnLysSerLysAsp 181
QY 624 AGCAGTACAGCTGCTGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 683
    |||||
DB 182 SerThrTyrSerMetSerSerThrLeuThrLeuThrLysAspLysLysLysLysLys 201
QY 684 GTTACGCTCCGCAAGTCAACCCATCAAGGCGCTAGCTCGCCGCTCAAAAGCTTCAAC 743
    |||||
DB 202 SerTyrThrCysGlnAlaThrHisLysTrpThrSerThrProLeuValLysSerPheAsn 221
QY 744 AGGAGAGAGTGT 755
    |||||
DB 222 ArgAsnGlnLys 225

```

## RESULT 11

```

S68241
I: kappa chain V region (Mab13-1) - mouse (fragment)
N: Alternate names: Immunoglobulin light chain
C: Species: Mus musculus (house mouse)
C: Date: 24-Aug-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C: Accession: S68241; S68214
R: Takagi, M.; Kohda, K.; Hamuro, T.; Harada, A.; Yamaguchi, H.; Kamachi, M.; Imanaka,
submitted to the EMBL Data Library, March 1994
A: Description: Specific peroxidase activity by formation of an antibody L-chain-porphyr
A: Reference number: S68241
A: Accession: S68241
A: Molecule type: mRNA
A: Residues: 1-218 <TAK>
A: Cross-references: EMBL:D29670; NID:q473962; PIDN:BA06141.1; PID:q473963
R: Takagi, M.; Kohda, K.; Hamuro, T.; Harada, A.; Yamaguchi, H.; Kamachi, M.; Imanaka,
FEBS Lett. 375, 273-276, 1995
A: Title: Thermostable peroxidase activity with a recombinant antibody L chain-porphyr
A: Reference number: S68211; MUID:56085223; PMID:7498516
A: Accession: S68214
A: Status: nucleic acid sequence not shown
A: Molecule type: mRNA
A: Residues: NR,3-212 <TAW>
A: Cross-references: EMBL:D29670
C: Superfamily: Immunoglobulin V region; Immunoglobulin homology
C: Keywords: immunoglobulin

```

## Alignment Scores:

```

Pred. No.: 1.62e-39 Length: 218
Score: 717.00 Matches: 137
Percent Similarity: 77.06% Conservative: 31

```



DB: 2 Gaps: 1

US-09-806-276A-3 (1-962) x A31790 (1-220)

QY 114 GAAATGTGTGACAGTCTCCAGCACCCTGTCTTGTCTCCAGGGAGAACGCCACC 173  
 DB 1 AsplLevalMetThrInserProSerSerLeuThrValThrIaGlyGlnValThr 20  
 QY 174 CTCCTCTGACAGGCCCATCAGATGTT-----AGCACTACTTAGCC 215  
 DB 21 MetSerCysThrSerSerGlnSerLeuPheAsnSerGlyGlnValSerValThr 40  
 QY 216 TGTGACCAACAGAACCTCGCCAGGCTCCAGGCTCTCATCTATGATGATCAACAGG 275  
 DB 41 TTPlyrGlnGlnValProGlyGlnProProlValLeuIleTyrTrpAlaSerThrArg 60  
 QY 276 GCCACTGGACATCCACCAGCCAGCTCAGTGGAGTGGCTGGAGACACTTCACTCACC 335  
 DB 61 GlnSerGlyValProAspArgPheThrGlySerGlySerGlyThrAspPheThrLeuThr 80  
 QY 336 ATCAGCAGACTGGAGGCCGGAAGATGTGACACTTATCTGTCTCAGCAATTTTACTACT 395  
 DB 81 IleSerSerValGlnAlaGlnAspLeuAlaValTyrTrpCysGlnAsnAspTyrSerAsn 100  
 QY 396 CCGTACACTTTTGGCCAGGAGCCAGGCTGGAGATCAACAACTGTGCTGCACCATCT 455  
 DB 101 ProlLeuThrPheGlyGlyGlyThrLysLeuGlnLeuLysArgAlaAspAlaIaProThr 120  
 QY 456 GTCTTCATCTTCCGCCCATCTGATGACAGCTGGAATCTGGAACCTGCTCTGTGTGTGC 515  
 DB 121 ValSerIlePheProProlSerSerGlnGlnLeuThrSerGlyGlyAlaSerValValCys 140  
 QY 516 CTGCTGATATCTCTATCCAGAGGCCCAACTAGTGAAGTGATGTAACGCCCTC 575  
 DB 141 PheLeuAsnAsnProPheTyrProlValAspIleAsnValIleStrpLysIleAspGlySerGln 160  
 QY 576 CAATCGGGTAACCTCCAGAGAGTGTCTCAGAGCAGACAGCAAGACACACCTACAGC 635  
 DB 161 ArgGlnAsnGlyValLeuAsnSerTrpThrAspGlnAspSerLysAspSerThrTyrSer 180  
 QY 636 CTCAGCAGCACCCTGAGCGCTGAGCAAAAGCAGACTACAGAAACAAAGCTTACGCTGC 655  
 DB 181 MetSerSerThrLeuThrLeuThrLysAspGlyIleValArgHisAsnSerTyrThrCys 200  
 QY 696 GAAGTCACCCATCAGGGCGCTGAGCTGCCGCTCAACAGAGCTTCAACAGGGAGAGTGT 755  
 DB 201 GlnAlaThrHisLysThrSerThrSerProIleValLysSerPheAsnArgAsnGlnCys 220

#### RESULT 14

monoclonal antibody 13-1 light chain - mouse

C:Species: Mus musculus (house mouse)

C:Date: 04-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 21-Jan-2000

C:Accession: J05810

R:Kashih, S.; Kato, K.; Torizawa, T.; Dojmae, N.; Yamaguchi, H.; Kamachi, M.; Harada, A.

Biochem. Biophys. Res. Commun. 240, 566-572, 1997

A:Title: Structural characterization of mouse monoclonal antibody 13-1 against a porphyrin

A:Reference number: J05810; MUID:96063277; PMID:9398605

A:Accession: J05810

A:Molecule type: protein

A:Residues: 1-218 <AKA>

C:Comment: This catalytic antibody has peroxidase oxidase. It is directed against a porphyrin

C:Superfamily: immunoglobulin V region; immunoglobulin homology

F:16-94/Domain: immunoglobulin homology <IMM>

#### Alignment Scores:

Pred. No.: 4,01e-39 Length: 218  
 Score: 711.00 Matches: 137  
 Percent Similarity: 76.50% Conservative: 29  
 Best Local Similarity: 63.13% Mismatches: 47  
 Query Match: 40.77% Indels: 4  
 DB: 2 Gaps: 1

US-09-806-276A-3 (1-962) x J05810 (1-218)

QY 117 ATTGTGTGACAGTCTCCAGCACCCTGTCTTGTCTCCAGGGAGAACGCCACCCTC 176  
 DB 2 IleValLeuThrGlnSerProAlaSerLeuAlaValSerLeuGlyGlnArgAlaThrIle 21  
 QY 177 TCTGAGAGGCGCAGTCTGAGTGTAGCAGC-----TACTTACCTGTAGACAA 224  
 DB 22 SerCysArgAlaSerLysSerValSerAlaSerGlyTyrIleTyrMetHisTrpTyrGln 41  
 QY 225 CAGAAACCTGGCCAGGCTCCAGGCTCTCATCTATGATGATCAACAGGAGGCCAGTGC 284  
 DB 42 GlnLysProGlyGlnProProlValLeuIleSerLeuAlaSerAsnLeuGlnIleSerGly 61  
 QY 285 ATCCACCCAGGTTTACGTGAGTGGCTGTGGGAGACACTTCACTTCCATCAGCAGA 344  
 DB 62 ValProAlaArgPheSerGlySerLysSerGlyThrAspPheThrLeuAsnIleHisPro 81  
 QY 345 CTGAGACCCAGAGATGTGACACTTATCTGTCTCAGCAATATTACTACTTCCGTACACT 404  
 DB 82 ValGlnGlnGlnAlaAspValAlaThrTyrTrpCysGlnHisSerArgGlnLeuProLeuThr 101  
 QY 405 TTTGGCCAGGAGCAGGCTGGAGATCAACAGCACTGTGCTGCACCATCTGTCTTCAATC 464  
 DB 102 PheGlyAlaGlyThrLysLeuGlnLeuLysArgAlaAspAlaIaProThrValSerIle 121  
 QY 465 TTTCCGCCCATCTGATGACAGTGAATCTGGAACCTGCTGTGTGTGTGTGTGTAT 524  
 DB 122 PheProProlSerSerGlnGlnLeuThrSerGlyGlyAlaSerValValCysPheLeuAsn 141  
 QY 525 AACTTCTATCCAGAGAGGCCCAACTAGTGAAGTGATGTAACGCCCTTCAATCGGGT 584  
 DB 142 AsnPheTyrProlValAspIleAsnValIleStrpLysIleAspGlySerGlnArgGlnAsn 161  
 QY 585 AACTCCAGAGAGTGTCTCAGAGCAGACAGCAAGACAGCAGCAGCTTACAGCTCAGCAGC 644  
 DB 162 GlnValLeuAsnSerTrpThrAspGlnAspSerLysAspSerThrTyrSerMetSerSer 181  
 QY 645 ACCCTGAGCTGAGCAACAGCAGTACAGCAAGCAAGTCTACGCTGGAGAGTCAAC 704  
 DB 182 ThrLeuThrLeuThrLysAspGlyIleValArgHisAsnSerTyrThrCysGlnAlaThr 201  
 QY 705 CATCAGGCGCTGAGCTGCCGCTCAACAGAGCTTCAACAGGGAGAGTGT 755  
 DB 202 HisLysThrSerThrSerProIleValLysSerPheAsnArgAsnGlnCys 218

#### RESULT 15

Ig kappa chain (Mab03-1) - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 29-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 11-Jan-2000

C:Accession: S68212

R:Yakagi, M.; Konda, K.; Hamuro, T.; Harada, A.; Yamaguchi, H.; Imanaka, F.

FEBS Lett. 375, 273-276, 1995

A:Title: The most stable peroxidase activity with a recombinant antibody L chain-porphyrin

A:Reference number: S68211; MUID:96085223; PMID:7498516

A:Accession: S68212

A:Molecule type: nucleic acid sequence not shown

A:Residues: 1-214 <YAK>

C:Cross-references: EMBL:D29668

C:Superfamily: immunoglobulin V region; immunoglobulin homology

#### Alignment Scores:

Pred. No.: 8.54e-39 Length: 214  
 Score: 706.00 Matches: 129  
 Percent Similarity: 79.44% Conservative: 41  
 Best Local Similarity: 60.28% Mismatches: 38  
 Query Match: 40.48% Indels: 6  
 DB: 2 Gaps: 1

US-09-806-276A-3 (1-962) x S68212 (1-214)



